THE UNIVERSITY OF CHICAGO

DETERMINANTS OF THE m^6A EPITRANSCRIPTOME

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

COMMITTEE ON IMMUNOLOGY

BY

P. CODY HE

CHICAGO, ILLINOIS

DECEMBER 2021

To my family.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

There are many people whom I would like to thank for their contributions to the work described in this thesis. First, I would like to express my gratitude to my PhD advisor, Chuan He. I am grateful that, even in my earliest days as a graduate student, he entrusted me with the freedom to pursue ambitious goals and tackle hard problems, even when the prospects for success were not necessarily clear. Without his tremendous support and mentorship, this work would not have been possible. Throughout my years in the lab, I felt that Chuan treated me as a partner in the research process and respected my ideas, which empowered me to take ownership of my work. Throughout the highs and lows of my thesis work, I found his support to be unwavering. He was always generous with his time, and his tireless work ethic and passion for his craft were inspiring. I am thankful for the lessons he has taught me, which have fostered my growth as a scientist and will undoubtedly serve me well in the future.

I would like to thank my committee members: Marcus Clark, Erin Adams, Tao Pan, Marcelo Nobrega and Mengjie Chen. Every committee member played an integral role in the completion of this thesis, with each making their own unique contribution; I could not have assembled a better group. I would especially like to thank Mengjie for our regular meetings to discuss the bioinformatic aspects of my research. I am grateful to have had such a group of talented, wise, and generous individuals guide me along this journey.

I would like to thank all members of the He Lab, past and present, for their support throughout my years as a PhD student. I feel blessed to have been able to work in such a highly collegial environment, where all lab members support each other with generosity and grace. In particular, I would like to thank Bryan Harada, who has generously given countless hours of his time over the years to discuss my research and give his thoughtful perspectives. I would also like

to recognize the contributions of Scott Zhang and Ruitu Lyu, who gave valuable input on many bioinformatic analyses. I want to express my gratitude to Hailing Shi, Jiangbo Wei, Tong Wu, Phil Hsu, Jun Liu, Qili Fei, Hui-Lung Sun and Kathy Liu, who helped train me early on in my graduate years. I would also like to thank Allen Zhu and Caraline Sepich-Poore, my fellow MD/PhD students in the lab, for their camaraderie and their generosity.

Next, I would like to acknowledge individuals outside of our laboratory who provided valuable input and assisted me with various aspects of my research; Jon Staley, Amelia Joslin, Shuai Wang, Pieter Faber and his colleagues at the University of Chicago Genomics Core, and Bill Buikema and his colleagues at the DNA Sequencing and Genotyping Facility. Thank you to Nancy Schwartz and Laurie Risner in the GDDTP MD/PhD program, the Committee on Immunology, and the Pritzker School of Medicine for supporting my graduate work. I would also like to thank the many friends I have met through these programs, especially members of the Pritzker entering class of 2016, who have filled my life with joy and meaning these past five years.

Finally, I would like to thank my family. I would like to thank my mom and dad, Ke Dong and Sheng Yang He. You both are incredible scientists, and even better parents. Growing up, your dinner table conversations captured my young scientific imagination and inspired me to try to make my own contributions to science one day. I'm grateful for your unconditional love and support and having you behind me, cheering me on every step of the way, no matter what path my life takes. Because of you, my dreams are limitless. I would like to thank my wife Jessi, who brings love and joy to my life every day. You have been there every step of the way, commiserating with the lows and celebrating the highs. Thank you for being my partner in this journey and believing in me and supporting my dreams; you inspire me to be the best that I can be.

ABSTRACT

N^6–methyladenosine (m^{6}A) is the most abundant mRNA modification and is crucial for gene expression regulation in numerous biological processes. In mammals, m^{6}A methylation is specifically deposited on a subset of mRNAs and is strongly enriched within long internal exons and near stop codons. This specificity is central to m^{6}A-mediated gene expression regulation, as m^{6}A reader proteins bind m^{6}A methylated regions of transcripts and exert downstream effects on gene expression. However, the basis of this specificity is poorly understood. Here, we develop a Massively Parallel assay for m^{6}A (MPm^{6}A) to systematically elucidate determinants of mRNA m^{6}A specificity. Unexpectedly, we discover that m^{6}A specificity is globally regulated by m^{6}A "suppressors" that prevent m^{6}A deposition in unmethylated transcriptome regions. We find that spliceosomes selectively suppress m^{6}A methylation in an exon length-dependent manner by depositing Exon Junction Complexes (EJC) that protect exon junction-proximal RNA from methylation. EJC suppression of m^{6}A underlies multiple global characteristics of m^{6}A specificity. EJC depletion causes aberrant methylation of thousands of mRNAs, resulting in widespread m^{6}A-mediated gene expression dysregulation. Altogether, we demonstrate that m^{6}A suppression by EJCs is a major mechanistic determinant of m^{6}A epitranscriptome specificity. Furthermore, we find that EJC-suppressed methylation sites co-localize with EJC-suppressed splice sites, suggesting that EJCs broadly suppress local mRNA accessibility to regulatory machineries by packaging proximal RNA.

INTRODUCTION[a]

RNA modifications and gene expression

The ability of organisms to read out the genetic instructions encoded in their DNA forms the basis for life. This genetic information is encoded in discrete units called genes, which comprise specific sequences of the four nucleotide monomers in DNA, with the nucleobases adenine (A), thymine (T), cytosine (C), and guanine (G). Genes specify the complex system of chemical processes that are required to sustain living systems. The process by which cells read out genetic information is called gene expression. During gene expression, the information contained in the DNA sequence is expressed into different chemical forms. First, the DNA sequence is copied into a complementary sequence of RNA through transcription by RNA polymerases. This RNA, called pre-messenger RNA (pre-mRNA), then undergoes a variety of processing steps, including capping, splicing, and polyadenylation, to produce mature messenger RNA (mRNA). The mRNA sequence is then translated from the 4-letter nucleotide alphabet of DNA and RNA into the 20-letter amino acid alphabet of proteins by ribosomes. Ribosomes collaborate with transfer RNAs (tRNA) that are bonded to amino acids to decode the mRNA sequence and synthesize proteins with the corresponding amino acid sequence. Proteins are the principal workhorses of the cell and are involved in virtually every cellular process. Proteins catalyze chemical reactions, act as structural scaffolds, organize subcellular structures, generate movement, sense signals, and perform many other essential biological processes. RNA, in addition to its function as an

[a] Parts of this section are reproduced or modified from He P.C., He, C. (2021) m^{6}A RNA methylation: from mechanisms to therapeutic potential. *The EMBO Journal*. e105977

information-carrying intermediate between DNA and protein, can also itself play certain functional roles through the catalysis of certain reactions and formation of protein-RNA complexes. This flow of genetic information from DNA to protein, termed the central dogma of molecular biology, is a fundamental feature of biological systems.

At any particular time, a typical human cell expresses only a subset of the ~20,000 genes in its genome. The set of genes that are actively expressed, and the abundance of the gene products, determines the function and identity of a particular cell. Thus, the ability of cells to robustly express their genes with high fidelity is crucial. The process of gene expression is finely regulated to ensure that the correct levels of gene products are produced. The rate and extent of gene expression is controlled at the transcriptional level, by factors such as transcription factors (TFs) and chromatin modifiers, at the post-transcriptional level, by factors such as RNA-binding proteins (RBPs) and RNA modification enzymes, and at the translational level, by factors such as kinases that phosphorylate translation initiation factors (Gehring et al., 2017; Jackson et al., 2010; Lee and Young, 2013). Different genes are subject to distinct modes of gene expression regulation by these factors due to variations in their sequences and architectural features. Differential expression of genes underlies diversity in cellular form and function across time and space. For example, different specialized cell types exhibit distinct gene expression programs that are tailored to their functions (Chen et al., 2018b). Changes in gene expression in response to outside stimuli enables changes in cellular behavior, such as in the immune response to invading pathogens (Carpenter et al., 2014). Finally, dysregulation of gene expression can lead to pathophysiological states, such as cancer (Bradner et al., 2017). Altogether, proper gene expression regulation is essential for diverse biological processes necessary for life.

Outside of the standard alphabet of nucleotides and amino acids that are used for gene expression, DNA, RNA, and protein also contain modified nucleotides or amino acids that differ in chemical structure from their canonical forms. Some of these modifications are the result of chemical damage from environmental sources, but many are due to targeted enzymatic modification of these biopolymers by dedicated cellular machineries (Chatterjee and Walker, 2017). These modifications impart modified DNA, RNA or protein with new properties, expanding the functional repertoire of these biopolymers. DNA and protein modifications have been extensively studied and have been shown to play essential roles in the control of gene expression and countless other cellular processes in physiological and pathophysiological settings (Allis and Jenuwein, 2016; Prabakaran et al., 2012; Suzuki and Bird, 2008). Like DNA, RNA contains four primary nucleotides, with the nucleobases adenine (A), uracil (U; instead of thymine), cytosine (C), and guanine (G). Cellular RNA additionally contains a diverse array of modified nucleotides (over 150 discovered so far) that are produced through enzymatic modification of the nucleobase or the ribose backbone (Boccaletto et al., 2018; Roundtree et al., 2017a). The modifications diversify the functionalities of these RNAs following their initial synthesis. Modifications can alter RNA structure by promoting or disrupting intramolecular interactions and can make RNAs more rigid or flexible (Ontiveros et al., 2019). They can also promote or disrupt RNA interactions with other biomolecules, such as RNA binding proteins that control gene expression or the antiviral immune response.

As a principal function of many cellular RNAs is to facilitate gene expression, it is perhaps unsurprising that many RNA modifications have been shown to impact gene expression. RNA comprises both a substrate that carries the genetic information, mRNA, as well as essential components of the gene expression machinery, including snRNAs that control splicing of pre-

mRNA, and rRNAs and tRNAs that enable the translation of mRNA into protein. All of these RNA species contain RNA modifications that are important for their gene expression functions (Roundtree et al., 2017a). mRNA modifications have been the focus of intensive study in recent years, largely driven by the development of new methods that map their presence on mRNA by next-generation sequencing technologies (Li et al., 2017b). These mRNA modifications include methylation of the nucleobase or ribose backbone, acetylation of the nucleobase, and isomerization of uridine to pseudouridine. Collectively, these mRNA modifications comprise the "epitranscriptome" (derived from the Greek prefix *epi-* (ἐπι- "over, outside of, around")), so termed because these modifications comprise a layer of features on top of the information encoded in the primary RNA sequence (Fig. 1) (Roundtree et al., 2017a). These mRNA modifications impact gene expression of modified transcripts through diverse mechanisms. In addition to the mRNA modifications that naturally occur on RNA, non-endogenous mRNA modifications can be incorporated onto synthetic mRNA for translational applications, such as the use of N^1-methylpseudouridine in COVID-19 mRNA vaccines to tune mRNA immunogenicity and stability (Nance and Meier, 2021). Here, we focus on the most prevalent naturally occurring mRNA modification in mammals, N^6-methyladenosine (m^6A).

m^6A methylation

N^6-methyladenosine (m^6A) is an adenosine derivative in RNA that plays wide-ranging roles in gene expression (He and He, 2021). m^6A is found in a variety of different species of RNA, including mRNA, rRNA, snRNA. The presence of m^6A in mRNA was first identified in the 1970s, and the methyltransferase complex that catalyzes m^6A mRNA methylation was purified in the

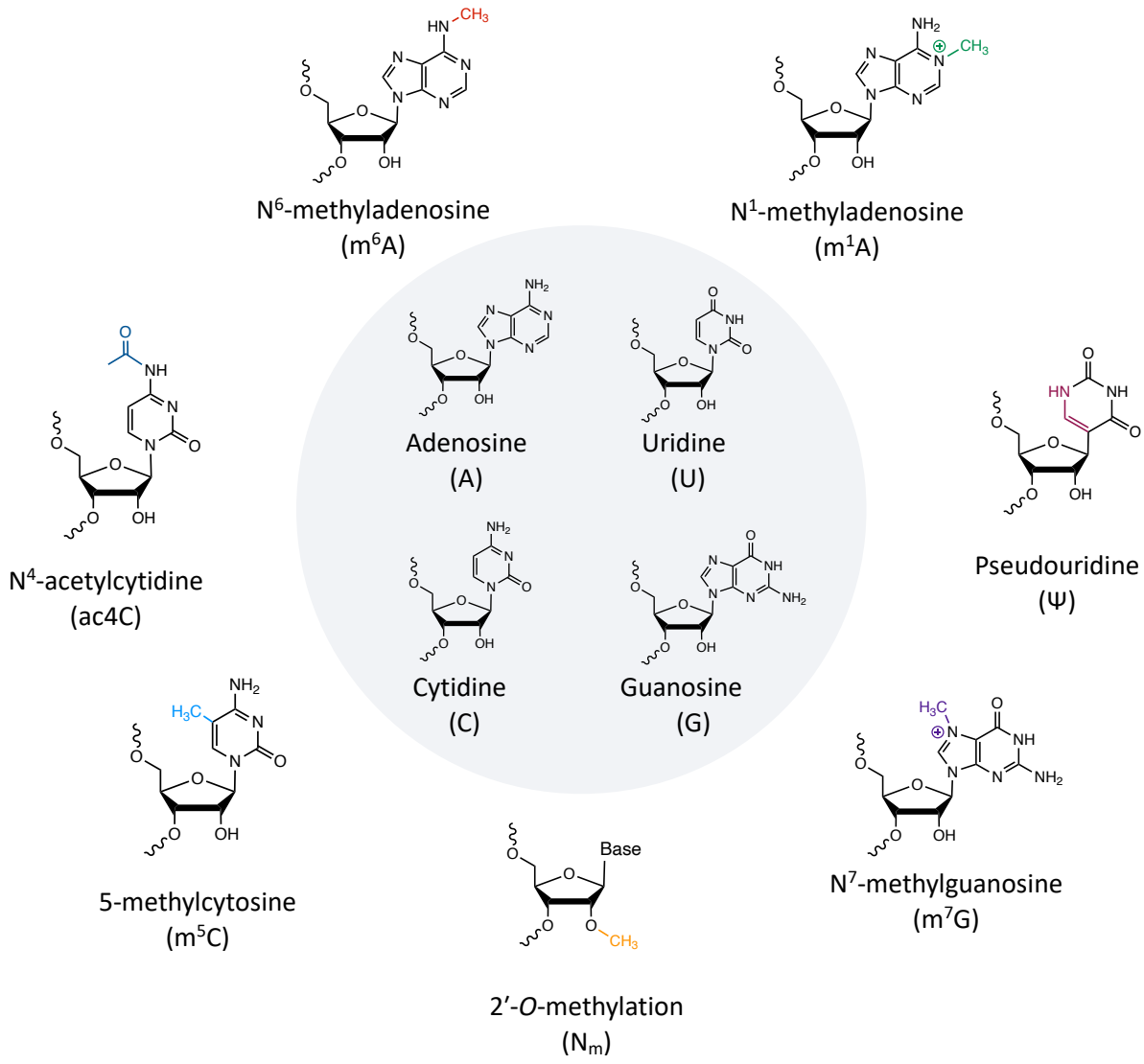

Figure 1. The epitranscriptome.

Chemical structures of the canonical four nucleosides in mRNA (center), and chemical structures of non-canonical modified nucleosides found in eukaryotic mRNA transcripts (outside). Colored regions indicate modifications.

1990s (Bokar et al., 1994, 1997; Desrosiers et al., 1974; Lavi and Shatkin, 1975; Perry et al., 1975; Wei and Moss, 1977; Wei et al., 1975). m^6A methylation is present in the RNA of many eukaryotes and is the most prevalent mRNA modification in mammals, where it is present at tens of thousands of sites across the transcriptome, at a frequency of 0.15-0.6% of all adenosines (Table 1) (Dominissini et al., 2012; Ke et al., 2015; Liu et al., 2014). m^6A is decoded by the ribosome as adenosine and therefore does not alter the genetic code. Instead, m^6A plays important roles in gene expression regulation, as m^6A methylation of mRNAs can alter the activities of cellular gene expression machineries on these modified RNAs.

A complement of m^6A pathway proteins have been identified and can be categorized into three groups according to their activities: "writers", which install m^6A methylation, "readers", which preferentially bind m^6A, and "erasers", which reverse m^6A methylation. The METTL3/METTL14 RNA methyltransferase complex is the predominant m^6A writer on mRNA. METTL3 is the catalytically active subunit, while METTL14 plays an essential structural role to facilitate catalysis(Liu et al., 2014; Wang et al., 2016a, 2016b). The METTL3/METTL14 heterodimer catalyzes the transfer of a methyl group from S-adenosylmethionine (SAM) to the N^6 position of adenosine within a DRACH (D = A, G or T; R = A or G; H = A, C or U) RNA consensus sequence. Orthologs of METTL3 and METTL14 have been identified in the sequenced genomes of many animals and plants, as well as some fungi (Bujnicki et al., 2002). METTL3/METTL14 associate with a large holocomplex approaching 1 megadalton in size composed of methyltransferase accessory subunits WTAP, VIRMA, RBM15A, RBM15B, ZC3H13 and HAKAI (Guo et al., 2018; Knuckles et al., 2018, 2018; Patil et al., 2016; Růžička et al., 2017; Wen et al., 2018; Yue et al., 2018) Some of these accessory subunits are required for the proper localization of METTL3/METTL14, and others have been suggested to influence the selective

deposition of m^{6}A by the methyltransferase complex. The METTL3/METTL14 methyltransferase complex typically exhibits a nuclear localization and deposits m^{6}A methylation onto mRNA transcripts, as well as other RNA Polymerase II (Pol II) transcribed RNAs, including long noncoding RNAs (lncRNA) and primary microRNAs (pri-miRNA) (Alarcón et al., 2015a; Ke et al., 2017; Knuckles et al., 2018; Slobodin et al., 2017a). m^{6}A methylation is specifically deposited on a subset of cellular mRNAs and is strongly enriched within certain transcript regions, such as in long internal exons and near stop codons (Dominissini et al., 2012; Ke et al., 2015; Meyer et al., 2012). m^{6}A is installed by METTL3/METTL14 in a substoichiometric manner; the percentage of cellular transcripts that are methylated at a particular site can vary from 2-3% to 80% or more (Garcia-Campos et al., 2019; Liu et al., 2013).

Deposition of m^{6}A by METTL3/METTL14 marks specific cellular transcripts for m^{6}A-mediated gene expression regulation (Figure 2). m^{6}A methylation modulates pre-mRNA processing, promotes mRNA nuclear export, controls mRNA localization, modulates mRNA stability, increases translation efficiency, and facilitates non-canonical translation initiation (Barbieri et al., 2017; Kasowitz et al., 2018; Liu et al., 2015, 2017; Meyer et al., 2015; Ries et al., 2019; Roundtree et al., 2017b; Slobodin et al., 2017a; Wang et al., 2014a, 2015; Xiao et al., 2016; Zhou et al., 2018, 2019a). Many of these effects are mediated by m^{6}A "reader" proteins that can selectively recognize m^{6}A and exert a regulatory function on m^{6}A-marked mRNA. m^{6}A readers include members of the YTH family, which bind m^{6}A with their eponymous YTH domains (Hsu et al., 2017; Roundtree et al., 2017b; Shi et al., 2017; Wang et al., 2014a, 2015; Xiao et al., 2016). The mammalian YTH family proteins consist of YTHDF1, YTHDF2, YTHDF3, YTHDC1 and YTHDC2. The mammalian YTHDF family of protein are cytosolic m^{6}A readers that regulate mRNA degradation and translation. YTHDF2, the first to have been characterized, promotes the

degradation of m^{6}A-modified RNAs for by multiple mechanisms (Wang et al., 2014a). YTHDF2 recruits the CCR4-NOT deadenylase complex to promote deadenylation and degradation of m^{6}A-marked mRNAs and promotes endoribonucleolytic cleavage of m^{6}A-marked mRNAs by the RNase P/MRP complex (Du et al., 2016; Park et al., 2019). The two other cytosolic YTHDF proteins, YTHDF1 and YTHDF3, have previously been reported to facilitate translation of their target methylated mRNAs. YTHDF1 promotes translation by recruiting the translation initiation factor eIF3 to bound mRNAs (Wang et al., 2015). YTHDF3 interacts with YTHDF1 and is thought to shuttle its bound transcripts to YTHDF1 to promote their translation (Shi et al., 2017). More recently, it has been proposed that YTHDF1, YTHDF2 and YTHDF3 act redundantly to decay m^{6}A methylated transcripts (Zaccara and Jaffrey, 2020). Further, YTHDF proteins have been proposed to control the subcellular localization of methylated RNAs by phase separating and partitioning into specific subcellular compartments (Ries et al., 2019).

In contrast to the cytoplasmic YTHDF proteins, YTHDC1 exhibits a nuclear localization and regulates gene expression steps that take place within the nucleus. YTHDC1 has been reported to modulate splicing, alternative cleavage and polyadenylation (APA), and nuclear export of m^{6}A methylated mRNAs (Kasowitz et al., 2018; Roundtree et al., 2017b; Xiao et al., 2016). YTHDC1 has also been reported to bind methylated non-coding chromosome-associated RNA transcripts and mediate their decay, which downregulates transcription of nearby mRNAs (Liu et al., 2020a). YTHDC2, the final member of the YTH family, is also a member of the DExH family of RNA helicases and contains multiple RNA-binding and helicase domains in addition to its YTH domain. YTHDC2 is an RNA-induced ATPase that exhibits 3’ to 5’ RNA helicase activity. YTHDC2 has been proposed to regulate translation through disruption of RNA secondary structures (Mao et al., 2019). It has also been implicated in mRNA degradation through its interaction with the 5’-3’

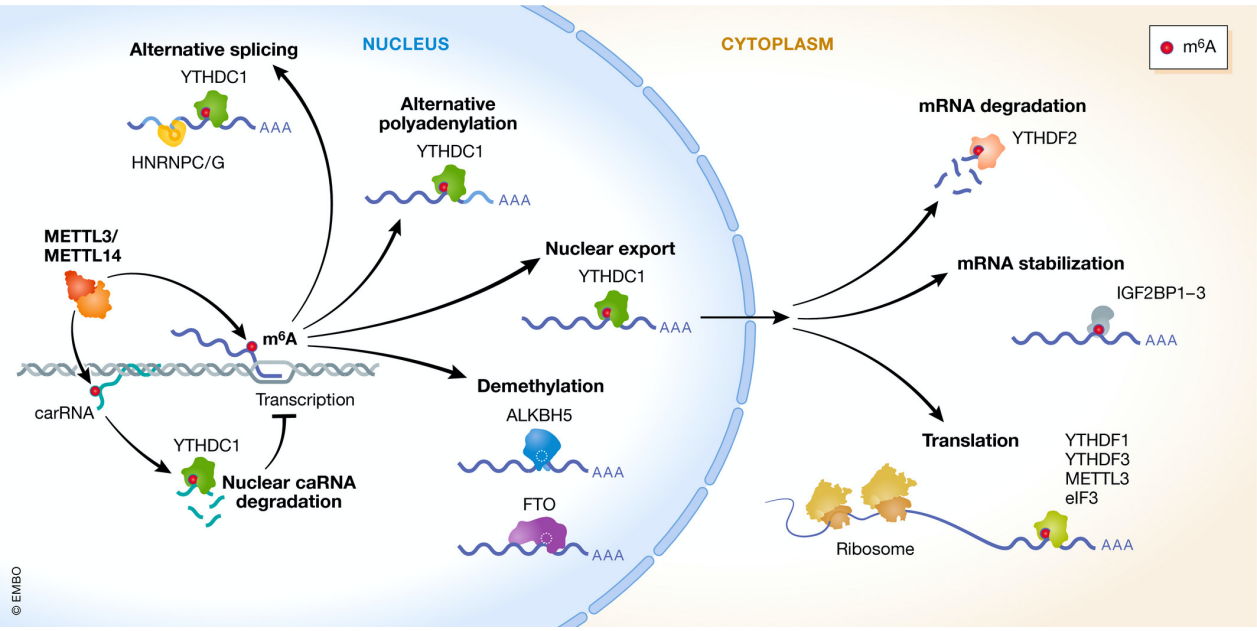

Figure 2. m^{6}A methylation is a multifaceted regulator of gene expression.

m^{6}A (red circle) regulates transcription, alternative splicing, alternative polyadenylation, nuclear export, cap-dependent and cap-independent translation, mRNA degradation and mRNA stabilization. A diverse set of m^{6}A reader proteins that preferentially bind m^{6}A methylated regions mediate these multifaceted effects on gene expression. Note that for clarity, nuclear processes are shown to occur after release of RNA from the polymerase, but some of the depicted nuclear processes may also occur co-transcriptionally. Figure originally published in (He and He, 2021).

exonuclease XRN1, though a direct effect on mRNA decay has not yet been demonstrated (Hsu et al., 2017; Kretschmer et al., 2018; Wojtas et al., 2017).

Deposition of m^6A on mRNA can also alter local RNA secondary structure and modulate the accessibility of proximal sequences to RNA-binding proteins that regulate gene expression. m^6A can exert stabilizing or destabilizing effects on secondary structure, depending on its sequence context, with destabilizing effects appearing more predominant transcriptome-wide (Kierzek and Kierzek, 2003; Roost et al., 2015; Spitale et al., 2015; Sun et al., 2019). m^6A methylation at "m^6A-switches", regions in which m^6A alters the structure of proximal RNA sequences to increase local accessibility to RBPs, facilitates binding of pre-mRNA processing factors HNRNPC and HNRNPG to m^6A methylated RNA (Liu et al., 2015, 2017; Zhou et al., 2019a). Several other RBPs that preferentially bind m^6A methylated regions have also been identified, but beyond the YTH family and HNRNPC/G, how these proteins selectively bind m^6A methylated regions has not been conclusively elucidated (Ae et al., 2017; Alarcón et al., 2015b; Edupuganti et al., 2017; Wu et al., 2018a). More recently, it has been suggested that m^6A may also regulate gene expression by disfavoring the binding of certain RBPs to mRNA, as has been reported for G3BP1 and LIN28A (Edupuganti et al., 2017; Sun et al., 2019). The wide array of mRNA processing factors whose activities are affected by m^6A methylation enable m^6A to regulate many of the diverse molecular processes involved in gene expression.

An important aspect of m^6A methylation is its reversibility, allowing for regulation of m^6A levels following initial deposition. Two RNA m^6A demethylases have been identified, FTO and ALKBH5. Both are iron- and α-ketoglutarate-dependent dioxygenases that mediate oxidative demethylation of m^6A. It was first reported that FTO demethylates m^6A on polyadenylated RNA, which was subsequently confirmed by several studies. Since these initial studies, other substrates

have since been reported; FTO has also been shown to demethylate m^6A and m^6A_m on mRNA, m^6A_m on U2 snRNA and m^1A on tRNA (Mauer et al., 2017, 2019; Wei et al., 2018). The substrate preferences for FTO appear to vary based on its subcellular localization -- the demethylation activity of FTO on m^6A is more pronounced in the nucleus, while its demethylation activity on m^6A_m is more pronounced in the cytoplasm (Wei et al., 2018). FTO has been reported to play roles in several steps of gene expression, from pre-mRNA processing steps, such as alternative splicing and alternative polyadenylation, to translation (Yu et al., 2018). ALKBH5 has a narrower substrate preference, appearing to be specific for m^6A. ALKBH5 has been reported to play important roles in splicing and nuclear export (Tang et al., 2018; Zheng et al., 2013).

Role of m^6A in physiological and pathophysiological processes

Given its wide-ranging roles in the regulation of gene expression, it is perhaps unsurprising that m^6A is essential for numerous physiological processes. Depletion of METTL3 homologs leads to defective meiosis in yeast, and developmental arrest in flies and plants (Agarwala et al., 2012; Bodi et al., 2012; Hongay and Orr-Weaver, 2011; Schwartz et al., 2014; Zhong et al., 2008). Genetic knockout of either *Mettl3* or *Mettl14* is developmentally lethal in mice, with embryos failing to thrive at around E5.5 (Batista et al., 2014; Geula et al., 2015). m^6A is present in the mRNA of all human and mouse tissues studied so far (Liu et al., 2020b; Xiao et al., 2019). Analysis of tissue-specific knockout mouse models of *Mettl3* and *Mettl14* have revealed essential roles for m^6A in brain development and function, cardiac homeostasis, immune system development and function, spermatogenesis, and skeletal function (Dorn et al., 2019; Li et al., 2017a; Lin et al., 2017; Rubio et al., 2018; Wang et al., 2018a, 2019, 2018b; Winkler et al., 2019; Wu et al., 2018b; Xu et al., 2020; Yoon et al., 2017). Moreover, knockout of either *Mettl3* or

Mettl14 severely blocks or delays differentiation in numerous stem cell or progenitor cell systems, including embryonic stem cells, embryonic neuronal stem cells (Wang et al., 2018b; Yoon et al., 2017), hematopoietic stem cells (Cheng et al., 2019; Lee et al., 2019; Vu et al., 2017; Weng et al., 2018; Zhang et al., 2017), naïve T cells (Li et al., 2017a), and bone marrow mesenchymal stem cells (Wu et al., 2018b).

Studies that implicate m^6A in the pathogenesis of a variety of diseases have emerged at a rapid pace in recent years. The role of m^6A in disease has been most extensively studied in the context of cancer, and important roles for METTL3/METTL14 in a variety of cancer types have been reported. Analysis of somatic mutations in cancer genomes from TCGA exome sequencing data revealed METTL3 and METTL14 as potential tumor suppressor genes in bladder and endometrial cancer, respectively (Zhao et al., 2019). Loss of METTL14 increases the proliferation and tumorigenicity of endometrial cancer cells by altering the mRNA stability and translation of AKT pathway regulators, resulting in increased AKT signaling (Liu et al., 2018). m^6A has also been reported to play tumor suppressor roles in renal cell carcinoma (Li et al., 2017b). Conversely, METTL3 and METTL14 have also been reported to act as oncogenes in other cancer types. METTL3 and METTL14 promote Acute Myeloid Leukemia (AML) oncogenesis by increasing the mRNA stability and translation of m^6A methylated oncogenes such as MYC (Barbieri et al., 2017; Vu et al., 2017; Weng et al., 2018). METTL3/METTL14 have also been found to play oncogenic roles in glioblastoma (Cui et al., 2017). In other cases, such as in hepatocellular carcinoma (HCC), the role of METTL3 and METTL14 is unclear due to contrasting findings regarding whether the m^6A methyltransferase complex promotes or inhibits cancer pathogenesis (Chen et al., 2018a; Ma et al., 2017). Due to the oncogenic roles of m^6A in certain cancer types, METTL3/METTL14 have been suggested to be promising drug targets. A small molecule catalytic

inhibitor of METTL3 has recently been reported to demonstrate efficacy in preclinical models of AML(Yankova et al., 2021). Several biotechnology companies have stated a primary focus on developing small molecule inhibitors for METTL3/METTL14 for testing in oncological indications, with Phase 1 trials planned for 2021-2022 (Cully, 2019). Overall, m^6A has emerged as a key regulator of numerous important biological processes in normal physiology and in disease.

Regulation of m^6A deposition on mRNA

A notable characteristic of the METTL3/METTL14 methyltransferase is the exquisite specificity with which it deposits m^6A on the transcriptome. The consensus motif for METTL3/METTL14 is a commonly occurring, degenerate DRACH consensus sequence. However, only a small fraction (~5%) of DRACH motifs in the transcriptome are selected for methylation (Table 1). Further, there is a marked regional bias in the distribution of m^6A across the transcriptome. m^6A is found on mRNAs of a subset of genes (estimated at 25-60%) and is strongly enriched within certain transcript regions, such as in long internal exons, stop codons and the beginning of last exons(Figure 3A) (Dominissini et al., 2012; Ke et al., 2015; Meyer et al., 2012). Internal exons that are >200 nt comprise only ~15% of all internal exons, but internal exons >200 nt contain ~80% of all m^6A sites present within internal exons (Ke et al., 2017). Additionally, transcript isoforms that use proximal APA sites are preferentially methylated compared to longer isoforms that use distal APA sites (Molinie et al., 2016).

This specificity is central to m^6A-mediated gene regulation, as it marks specific cellular transcripts for regulation by m^6A reader proteins. m^6A readers bind these m^6A-marked transcripts and exert downstream effects on gene expression, which is important for gene regulation in numerous physiological and pathophysiological processes. For example, in mouse embryonic stem

cells (mESCs), 80% of genes that regulate naïve pluripotency exhibit m^{6}A methylation on their transcripts. This is thought to facilitate clearance of these transcripts during the termination of pluripotency to enable mESC differentiation (Geula et al., 2015). Further, the specific position at which an mRNA is m^{6}A methylated may also be relevant for gene expression, since the nature of m^{6}A-mediated regulation appears to differ depending on the location of m^{6}A within the transcript (Barbieri et al., 2017; Choi et al., 2016; Mao et al., 2019; Meyer et al., 2015; Slobodin et al., 2017; Wang et al., 2014a, 2015; Zhou et al., 2018).

Given the relevance of specific m^{6}A deposition for gene expression regulation, several recent studies have focused on elucidating the mechanisms that control m^{6}A deposition. The transcription factors CEBPZ and SMAD2/3 promote methylation of specific transcripts by recruiting the methyltransferase complex to their targeted genomic loci(Barbieri et al., 2017; Bertero et al., 2018a). TARBP2, an RNA binding protein, can recruit the methyltransferase complex to its bound transcripts (Fish et al., 2019). Slow transcription rates have been proposed to promote m^{6}A deposition by enhancing co-transcriptional recruitment of the methyltransferase complex to its RNA substrate (Slobodin et al., 2017a). The histone H3K36me3 modification has also been proposed to recruit the methyltransferase complex to chromatin in order to promote methylation of nascent RNAs (Huang et al., 2019). Collectively, these studies have shed light on how diverse biological inputs can modulate m^{6}A deposition and lead to differential downstream effects on gene expression and cellular phenotypes (Figure 3B).

While progress has been made in elucidating new mechanisms regulating m^{6}A deposition, fundamental questions regarding m^{6}A deposition remain unanswered. The global enrichment of m^{6}A at specific transcriptomic features such as long internal exons and near stop codons, as well as selective deposition of m^{6}A on a large portion of methylated transcripts, are not explained by

Table 1. Statistics for m^6A in mRNA

Quantity	Value	References	Notes
Ratio of m^6A to A in polyadenylated RNA.	.15% - .6% (LC-MS/MS)	(Du et al., 2018; Lin et al., 2017; Liu et al., 2014, 2018, 2020b; Wang et al., 2014b; Wei et al., 2018)	m^6A to A ratios in polyadenylated RNA are heavily influenced by the highly expressed genes in the transcriptome. In humans, a small fraction of all expressed genes (10 to 1000 genes, depending on the tissue type), make up the majority of the mRNA pool (Ramsköld et al., 2009).
Percentage of transcripts in the transcriptome that contain m^6A	~25% - ~60%. (m^6A-seq)	(Dominissini et al., 2012, unpublished)	Dominissini et al. identified m^6A in transcripts of ~7000 genes. The 25% figure is derived from dividing 7000 from the estimated number of total human genes, ~20,000-25,000. However, this is likely an underestimate. First, all genes are not expressed in any given cell type. Second, m^6A in relatively lowly-expressed genes escapes detection due to low sequence coverage (increasing sequencing depth increases the number of m^6A peaks identified). By restricting the analysis to a set of transcripts above an expression cutoff set to the lower quartile of expression for detected methylated transcripts, the percentage of transcripts containing at least one m^6A peak is higher, ~ 60% (unpublished). It is important to note here that this percentage reflects the number of genes whose transcripts exhibit m^6A, not the stoichiometry of m^6A in the transcripts of any given gene.
Average number of m^6A per transcript	~3 m^6A residues per average mRNA transcript (LC-MS/MS), ~ 1.7 m^6A-seq peaks per transcript (m^6A-seq)	(Dominissini et al., 2012; Perry et al., 1975)	It is important to consider that these are averages. Many transcripts are unmethylated and many have significantly more m^6A peaks than the average.

Percentage of DRACH motifs that are methylated	~5%	(Dominissini et al., 2012)	Dominissini et al. estimate that a prevalence of ~ 1 m^6A peak every 2000 bp. The number of DRACH motifs found in 2000 bp of sequence is theoretically (3/4)(2/4)(1/4)(1/4)(2/4) * 2000 = 23.4. Thus, the percentage is estimated at 1/23.4 ~ 5%. The actual percentage may be several-fold higher or lower, given that m^6A peaks may contain several m^6A sites and the true frequency of DRACH motifs likely diverges from the theoretical estimate.

known pathways. Additionally, while certain mechanisms that involve regulation of METTL3/METTL14 activity by trans-acting factors have been characterized, the relative importance of these mechanisms to shaping the global m^6A landscape in comparison to other factors, such as intrinsic preference of METTL3/METTL14 for certain RNA sequences or structures, is unclear.

The degree to which intrinsic vs. extrinsic factors determine m^6A levels has been the subject of debate: to what extent are m^6A levels determined by the intrinsic preference of the METTL3/METTL14 methyltransferase for certain nucleotide sequences vs. the contribution of extrinsic regulation of methyltransferase complex activity by external factors such as RBPs, TFs, RNA polymerase, or other yet to be discovered factors? This question is important due to its potential implications for m^6A as a regulatory system. If intrinsic determinants are dominant, this would suggest that m^6A deposition is not highly regulated and would generally operate independently of other cellular processes. It would suggest that global changes in m^6A methylation could occur if enzyme levels or other reaction variables change, but relative changes of m^6A levels at different sites would be limited. Conversely, if extrinsic determinants dominate, this would indicate a more dynamic role for m^6A, in which m^6A levels may react to changes in the activity of external factors in different cellular states. It would moreover imply that m^6A deposition is interwoven with other cellular processes, allowing for subsets of sites to gain or lose methylation in response to specific cellular events. The involvement of extrinsic determinants is supported by the existence of many studies that report differential methylation of significant numbers of m^6A peaks in different cellular contexts (Aguilo et al., 2015; Engel et al., 2018; Hesser et al., 2018; Lichinchi et al., 2016b, 2016a; Liu et al., 2020b; Tan et al., 2018; Xiao et al., 2019; Yu et al., 2018; Zhou et al., 2015). Additionally, several mechanisms through which trans-acting factors modulate

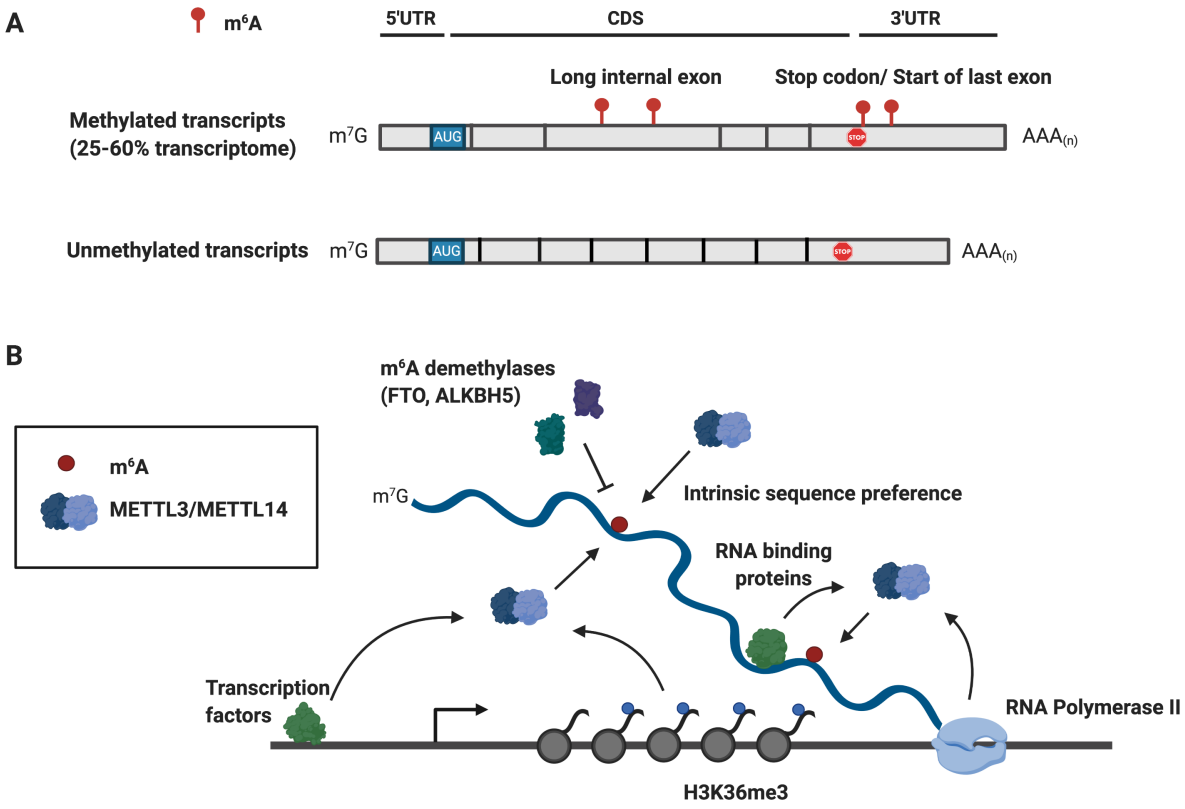

Figure 3. Selective m^{6}A methylation of the transcriptome.

A, Schematic representing the distribution of m^{6}A in the mammalian transcriptome. A subset of transcripts are methylated, while another subset are not. On methylated transcripts, m^{6}A is enriched in unusually long internal exons and near stop codons/start of last exons. **B,** Schematic depicting control mechanisms governing m^{6}A installation. m^{6}A deposition is regulated by intrinsic factors, such as the preference of the METTL3/METTL14 methyltransferase for specific RNA sequences. m^{6}A deposition is also regulated by extrinsic factors; transcription factors, RNA-binding proteins, RNA polymerase II, and the H3K36me3 histone modification have been reported to recruit the METTL3/METTL14 methyltransferase to mRNAs to promote methylation. m^{6}A demethylases FTO and ALKBH5 can also tune m^{6}A levels at a subset of sites.

methylation deposition have been characterized, as described above. However, some have questioned the validity of some of these conclusions, reporting limited reproducibility of differential methylation upon independent re-analysis (McIntyre et al., 2020). Further, other reports have concluded that intrinsic determinants are predominant and that m^6A levels are globally "hard-coded" in cis by the local sequence surrounding the m^6A site (Garcia-Campos et al., 2019; Schwartz et al., 2014). Part of this apparent discrepancy in the predominance of intrinsic or extrinsic determinants may be due to differences in m^6A regulation in different organisms. Most studies characterizing the importance of extrinsic factors have been carried out in mammalian cells (Barbieri et al., 2017; Bertero et al., 2018a; Fish et al., 2019; Slobodin et al., 2017a). One study reporting that intrinsic determinants largely control m^6A levels primarily examined the ability of local sequence features to predict m^6A methylation levels in the budding yeast *S. cerevisiae* (Garcia-Campos et al., 2019). The authors found that in mammalian systems, the ability to predict m^6A methylation levels from local sequence features was diminished compared to methylation levels in *S. cerevisiae*. Part of this discrepancy may also arise from an assumption that local sequence variation only alters intrinsic determinants and not regulation by extrinsic factors, which may not necessarily hold true. However, as this question has not yet been systematically explored, the nature of global m^6A deposition remains largely enigmatic.

Despite recognition of the strong enrichment of m^6A within long exons and near stop codons since the first transcriptome-wide m^6A mapping studies, the mechanisms regulating m^6A methylation reported to date do not explain how m^6A is selectively deposited in this specific distribution. The m^6A methyltransferase accessory factor VIRMA appears to promote methylation of many sites in mRNA 3' untranslated regions (UTRs) and near stop codons, but the mechanistic basis for this effect is unclear (Yue et al., 2018). H3K36me3 directly recruits METTL14 to gene

bodies to promote m^{6}A deposition, but does not appear to confer enrichment of m^{6}A near stop codons, as this enrichment is still observed when H3K36me3 levels are reduced (Huang et al., 2019). Most reported mechanisms of m^{6}A deposition regulation involve models in which a trans-acting factor recruits the METTL3/METTL14 methyltransferase to RNAs to promote methylation at proximal sites. Whether mechanisms with different modes of action contribute to m^{6}A specificity has remained largely unexplored. The inability of currently known mechanisms to explain how m^{6}A is specifically enriched in certain transcriptome regions suggests that major mechanisms regulating m^{6}A deposition globally remain to be discovered.

Determinants of the m^{6}A epitranscriptome

It has been nearly a decade since the m^{6}A epitranscriptome was first comprehensively mapped, revealing the distinctive distribution of m^{6}A methylation across the transcriptome. Mechanisms by which m^{6}A methylation controls gene expression and physiological functions of m^{6}A methylation have since been characterized. In contrast, the determinants of m^{6}A methylation have remained poorly understood. Out of the many potential m^{6}A sites in the transcriptome, why are only a few selected for methylation? What is the relative importance of intrinsic vs extrinsic factors in determining the m^{6}A epitranscriptome? To what extent is m^{6}A deposition interwoven with other cellular processes? Work contained herein sheds light on these key questions.

MATERIALS AND METHODS

Cell lines

HeLa and HEK293T cell lines used in this study were grown in Dulbecco's modified Eagle's medium (Gibco, 11995040) and supplemented with 10% fetal bovine serum (FBS) (Gibco), and 1% penicillin–streptomycin (Gibco). TREx FLAG-Magoh and TREx FLAG-eIF4A3 cell lines were grown in Dulbecco's modified Eagle's medium (Gibco, 11965126) and supplemented with 10% fetal bovine serum (FBS) (Gibco), and 1% penicillin–streptomycin (Gibco). Cell lines were maintained at 37°C with 5% CO2. Cells were passaged at ~90% confluency every 2–3 days. Cells were tested for mycoplasma by PCR (Uphoff and Drexler, 2004). The identity of HeLa and HEK293T cell lines were authenticated with STR profiling (ATCC). HeLa *METTL3*$^{-/mut}$ cells were generated as described in (Yue et al., 2018). TREx FLAG-Magoh and TREx FLAG-eIF4A3 cell lines were gifts from Dr. Guramrit Singh and were generated as described in (Singh et al., 2012).

Antibodies

The primary antibodies were purchased from commercial sources, with information about the antibodies and their use in this study listed in Table 8.

Oligonucleotides

All individual oligonucleotides in this study were synthesized by IDT, with sequences listed in Table 8.

Molecular cloning

pAc-GFP-Hyg-C1 A1335T was generated from pAc-GFP-Hyg-C1 (Clontech) by site-directed mutagenesis using the QuikChange Site-directed mutagenesis kit (Agilent). This A to T substitution at position 1335 disrupts a DRACH motif predicted to be an m^6A methylation site (Zhou et al., 2016). The pTBG plasmid was a gift from Dr. Ligang Wu (Du et al., 2016). pBG was constructed by cloning the BG insert from pTBG into the pAc-GFP-HYG-C1 backbone. pBG Δ i1, i2, pBG Δ i1, and pBG Δ i2 and pBG Δss were constructed by cloning synthesized gBlocks (IDT) into the pAc-GFP-Hyg-C1 backbone with PCR amplification to add complementary sequences and assembly with NEBuilder HiFi DNA assembly master mix. *CRY1* suppressed m^6A site sequence with varying amounts of flanking sequence was cloned into BG plasmids by synthesis of oligonucleotide or Gblock (IDT), PCR amplification to add complementary sequences and assembly with AccI and BamHI-HF (NEB) digested plasmids with NEBuilder HiFi DNA assembly master mix. pcDNA3.1+ SMN1 Myc His was a gift from Dr. Gideon Dreyfuss. pcDNA3.1+ EV was generated by PCR amplification add complementary sequences and assembly with NEBuilder HiFi DNA assembly master mix. Cloning products were transformed into NEB Turbo Competent E. coli, High Efficiency (NEB). All cloning primers can be found in Table 8. Description of plasmids can be found in Table 8.

siRNA knockdown

Dharmacon ON-TARGET Plus siRNAs were used for siRNA knockdown, with sequences listed in found in Table 8. Cells were seeded into TC-treated culture dishes (Corning) 16–20 h before transfection to yield ~30% confluency on the day of transfection. siRNA was transfected at 40 nM concentration with Lipofectamine RNAiMAX (Invitrogen) per the manufacturer's protocol.

Cells were incubated with siRNA for 5 hours and then media was replaced with fresh media. Cells were collected 48 hours post-transfection. siRNA sequences can be found in Table 8.

Plasmid transfection

Cells were seeded into tissue culture-treated dishes (Corning) 16–20 h before transfection to yield ~80% confluency on the day of transfection. Plasmids were transfected using Lipofectamine 2000. For every μg of plasmid transfected, three μl of Lipofectamine 2000 were used. For MPm^{6}A library transfections, 10 μg of library was transfected into 80% confluent 10 cm plates of HeLa cells, and cells were collected 24 hours later. For pBG transfections, 5 μg of library was transfected into 80% confluent 10 cm plates of HeLa cells. Cells were incubated with plasmid/lipofectamine mixture for 4 hours and then media was replaced with fresh media. Cells were collected 24 hours post-transfection.

RNA isolation

For MPm^{6}A, m^{6}A-seq, m^{6}A-IP-RT-qPCR and SELECT, total RNA was isolated by adding TRIzol reagent directly to cell culture plates (Invitrogen), followed by chloroform extraction (5:1 Trizol:Chloroform), precipitation of RNA from the aqueous phase with isopropanol (2:1 original volume of Trizol:Isopropanol), washing of the RNA pellet with 80% ethanol, and then resuspension in RNase-free water. RNA concentration was measured by ultraviolet absorbance at 260 nm using a Nanodrop instrument. Polyadenylated RNA was purified from total RNA using the Dynabeads mRNA purification kit (Invitrogen), with the following modifications. Lysis Binding Buffer from the Dynabeads Direct mRNA purification kit (Invitrogen) was used instead of Binding Buffer, and the initial incubation of mRNA at 65°C was omitted. mRNA was eluted in

nuclease-free water and concentration was measured by ultraviolet absorbance at 260 nm using a Nanodrop instrument. RNA samples used for RT-qPCR were isolated by using a Direct-zol RNA miniprep kit (Zymo Research) with the optional on-column DNase-I digestion step.

m^{6}A-meRIP-seq

Polyadenylated RNA was isolated as described above. RNA was adjusted to 10 ng/µl in 100 µl H_2O, 2 µl of 100 pM GLuc m^{6}A methylated RNA (NEB) transcribed in the presence of 20% m^{6}ATP and 80% ATP and 2 µl of 100 pM CLuc (NEB) unmethylated RNA was spiked in, and RNA was sheared using a Bioruptor Ultrasonicator (Diagenode), with 30 cycles of 30 seconds on/off at 4°C to generate ~150nt fragments. 5% of RNA was saved as input and the remainder underwent m^{6}A-immunoprecipitation (m^{6}A-IP) using the EpiMark N6-Methyladenosine Enrichment Kit (NEB). SUPERase-in RNase inhibitor was added to all buffers for a concentration of 20 U/mL. Eluted RNA was purified using RNA Clean and Concentrator (Zymo Research). Library preparation of input and IP samples was performed with the Illumina Truseq Stranded mRNA Library Prep kit. Sequencing was carried out on an Illumina HiSeq4000 machine with SE 50 bp reads. HISAT2 (version 2.1.0) (Kim et al., 2015) was used to align the sequence reads to reference genome (hg38) with options -k 1 and --known-splicesite-infile generated by the hisat2_extract_splice_sites.py script with the UCSC hg38 GTF. samtools (version 1.7) was used to generate, sort and index bam files. The callPeakBinomial function and reportJointPeak from the MeRIPtools R package (version 0.1.0) was used to call m^{6}A peaks with options min_counts = 15, peak_cutoff_fdr = 0.05, peak_cutoff_oddRatio = 1.5, and joint_threshold = 2. RADAR R/Bioconductor package (version 0.2.1) (Zhang et al., 2019) was used to perform differential methylation analysis with options fragmentLength = 150, binSize =50, minCountsCutOff = 10,

FDR cutoff = .1, log2FC cutoff = 1. Significantly differentially methylated regions were annotated to the RefSeq database (hg38). ClusterProfiler (version 3.6) was used to carry out GO term enrichment analysis (Yu et al., 2012). Metagene profiles were constructed with MetaPlotR. The center position of peaks was used for plotting. If the center of the peak fell in an intron, the end position of the peak was used for plotting. UTRs were scaled to reflect the average size of UTRs in the transcriptome. Bedgraph files for the forward and reverse strand were created using deeptools (version 3.1.3) bamCoverage with options --binSize 10, --normalizeUsing CPM, --exactScaling, --filterRNAstrand forward/reverse. Bedgraph files were visualized using the UCSC genome browser (Kent et al., 2002). To visualize enrichment of input and m^{6}A IP read density surrounding splice sites, deeptools ComputeMatrix and Plotprofiles were used with splice sites obtained from the hisat2_extract_splice_sites.py script and hg38 GTF.

MPm^{6}A oligonucleotide library design

59,730 148-nt oligonucleotides were synthesized as a SurePrint Oligonucleotide Library (Agilent). These sequences consist of a 17 nucleotide (nt) 5' universal priming sequence (ACTGGCCGCTTCACTGC), 102 nt of genomic sequence, a 12 nt barcode sequence, a 17 nt 3' universal priming sequence (AGATCGGAAGAGCGTCG). The universal priming sequences were constant for all oligonucleotides and were used to clone the sequences into pAc-GFP-Hyg-C1 A1335T to create a plasmid library, and for Illumina NGS library construction. For "experimental" sequences, the 102 nt genomic sequence consists of an endogenously m^{6}A methylated site or endogenously unmethylated DRACH site, along with 50/51 nucleotides of the flanking genomic sequence for all experimental sequences. If possible, a sequence that does not introduce a stop codon when cloned in-frame was chosen. A minimum read count filter was

imposed at the position of the DRACH sequence to filter out DRACH sites that are potentially false-negative m^{6}A sites due to low read coverage. Sequences that did not fall completely within exons were filtered out. Each experimental sequence had a corresponding mutant sequence that served as a negative control to control for potential non-specific enrichment. All "mutant" sequences were identical to experimental sequences but with all RACs mutated to RTT. The 12 nt barcodes were generated using the DNABarcodes R package(Buschmann and Bystrykh, 2013). All barcodes that created RAC sequences or microRNA seed sequences(Agarwal et al., 2015) when assembled into the full oligonucleotide sequence or matched were filtered out.

6897 endogenously methylated sequences were chosen for use in the MPm^{6}A assay. The m^{6}A sites were sampled from m^{6}A sites taken from a published m^{6}A antibody-crosslinking based HeLa single-base resolution m^{6}A dataset (m^{6}A-CLIP)(Ke et al., 2017). To reduce potential false positive m^{6}A sites, we required that the m^{6}A site be present in the cytoplasmic, nuclear and chromatin-associated m^{6}A-CLIP datasets from Ke et al., 2017 and fall within HeLa m^{6}A-seq peaks.

3058 endogenously unmethylated sequences were chosen for use in the assay. To generate the endogenous sequences containing non-methylated DRACH sites, Ensembl HG38 transcriptome cDNA and ncRNA sequences were input into a random forest classifier that predicts m^{6}A sites based on sequence features (Zhou et al., 2016). The sequences were sampled from sites predicted to be methylated but which exhibited no cellular methylation in HeLa cells. We implemented a conservative filtering criteria for unmethylated sites, choosing sites that are at least 200 nt away from all m^{6}A-CLIP sites and at least 200 nt away from all HeLa m^{6}A-seq peaks. We implemented a minimum read coverage cutoff as an additional filter to eliminate potential false-negative m^{6}A

sites deriving from low expression levels. Additionally, we only chose sites that exhibited depletion of IP reads compared to input at that site (log_2Fold change < -1) to eliminate any potential lowly-methylated m^{6}A sites that fall just under the threshold for significance by m^{6}A-seq or m^{6}A-CLIP. Each experimental and mutant sequence were associated with three distinct barcodes to generate three internal replicates per biological replicate.

MPm^{6}A plasmid library construction

The oligonucleotide pool was amplified by emulsion PCR (Micellula DNA Emulsion & Purification (ePCR) Kit, ChimerX) with Herculase II Fusion DNA polymerase (Agilent) and primers that anneal to the 5' and 3' priming regions and add sequences needed for Gibson assembly cloning. Amplified DNA was purified using DNA Clean and Concentrator-5 (Zymo Research). pAcGFP1-Hyg-C1 A1335T was digested with restriction enzymes overnight at 37°C (BclI-HF(NEB) for 3'UTR MPm^{6}A, BstEII-HF(NEB) for CDS MPm^{6}A, Age1-HF (NEB) for 5'UTR MPm^{6}A), and then gel purified with Qiaquick Gel Extraction kit (Qiagen) to remove contaminating uncut plasmid. ePCR-amplified sequences were cloned by Gibson Assembly with NEBuilder HiFi DNA assembly master mix (NEB) to generate a library of reporter constructs. The assembly reaction was incubated for 1 hour at 50°C, then purified with DNA Clean and Concentrator-5. Restriction sites used for cloning were destroyed by successful assembly. To eliminate any residual uncut plasmid, DNA was isolated and treated with the original restriction enzyme used for cloning and Exonuclease V (RecBCD) (NEB). The DNA was purified using DNA Clean and Concentrator-5 and then 1 µl of eluted DNA was transformed via electroporation into 10 NEB10-beta tubes, using the high efficiency transformation protocol. Transformed cells were

spread on 10 15 cm LB agar petri dishes and grown at 37°C overnight. Dilutions were also plated to estimate library coverage. The following day, colonies were scraped into LB and then plasmid DNA was purified with two Zymopure Maxiprep (Zymo Research) columns. Library quality was checked by Sanger sequencing of the plasmid library as well as individual clones that were not pooled into the library.

Cellular MPm^{6}A experimental procedure

10 μg of the MPm^{6}A plasmid library was transfected into a 80% confluent 10 cm plate of HeLa cells for cellular transcription and m^{6}A methylation. 24 hours post-transfection, polyadenylated RNA was isolated as described above. Polyadenylated RNA was incubated with 2U of Turbo DNAse in a 50 μl reaction for 37°C for 1 hour to eliminate plasmid DNA. RNA was adjusted to 10 ng/μl in 100 μl H2O, 2 μl of 100 pM GLuc m^{6}A methylated RNA transcribed in the presence of 20% m^{6}ATP and 80% ATP and 2 μl of 100 pM CLuc unmethylated RNA was spiked in. 5% of RNA was saved as input and the remainder underwent m^{6}A-immunoprecipitation (m^{6}A-IP) using the EpiMark N6-Methyladenosine Enrichment Kit (NEB). cDNA was generated using Superscript IV (Thermo) and a site-specific RT primer (Universal Primer 2) that hybridizes to the 3' universal priming sequence. Next generation sequencing libraries were prepared from cDNA by PCR with NEBNext Ultra II Q5 master mix (NEB) using the recommended NGS library PCR cycling conditions. PCR primers that anneal to the 5' and 3' universal priming sequences and add Illumina adaptor sequences were used for library PCR. PCR amplification cycles were determined by qPCR by calculating the number of cycles required for ½ max fluorescence. To control for PCR bias across samples, all input libraries were amplified the same number of cycles, and all IP were

amplified the same number of cycles. Library quality was checked by Agilent Bioanalyzer. Libraries were sequenced on an Illumina HiSeq4000 with SE 50 bp reads.

In vitro MPm^{6}A experimental procedure

A T7 promoter sequence was added upstream of the TSS through emulsion PCR of the MPm^{6}A library with primers that anneal to the TSS and cleavage/polyadenylation site. The MPm^{6}A library was in vitro transcribed (IVT) using NEB HiScribe T7 High Yield RNA Synthesis Kit. RNA was incubated with 2U of Turbo DNAse in a 50 µl reaction for 37°C for 1 hour to eliminate plasmid DNA. The purified IVT library was incubated with recombinant METTL3/METTL14. The in vitro methyltransferase reaction was carried out in a 25 µl reaction containing the following components: 1.5 µg IVT RNA, 300 nM recombinant METTL3/METTL14 complex (Active Motif), 1.6 mM SAM, 80 mM KCl, 1.5 mM MgCl2, 0.2 U µl−1 SUPERase-IN RNase inhibitor (Invitrogen), 10 mM DTT, 4% glycerol and 20 mM Tris-HCl (pH 8.0). Mock treatments omitted recombinant METTL3/METTL14. For mock-treated RNA, all reaction components except for the recombinant METTL3/METTL14 complex were added. The reaction was incubated for 22 h at 16 °C; RNA was purified with a Monarch RNA cleanup kit (NEB). A portion of RNA was digested with nuclease P1 and alkaline phosphatase for LC-MS/MS detection to check m^{6}A methylation levels. The remaining RNA was immunoprecipitated with an anti-m^{6}A antibody and further processed as described in the cellular MPm^{6}A method.

MPm^{6}A computational procedure

Reads were mapped back to a reference genome containing the designed oligonucleotide sequences with HISAT2 (version 2.1.0), with option -k 1. Counts for each sequence were

calculated with samtools (version 1.7) idxstats and sequences with at least 10 counts for all barcoded experimental and mutant sequences were retained. IP/input ratios are calculated for each sequence, representing the m^6A enrichment level that is present on each sequence. To calculate an m^6A enrichment score for each experimental sequence, we took the average of the IP/input ratios of the three experimental sequence internal replicates and three mutant sequence internal replicates over three or four biological replicates. We then took the difference of the experimental and mutant negative control sequence to calculate the enrichment score. Statistical significance of enrichment was determined with the Wilcoxon Rank Sum test with the Benjamin-Hochberg procedure to control FDR.

Exon length analysis

For exon length analysis, first, internal and last exons were obtained from the TxDb.Hsapiens.UCSC.hg38.knownGene R/Bioconductor package (version 3.4.0). For exon length analysis, only m^6A peaks that unambiguously reside within first, internal or last exons were included for analysis. If an m^6A peak overlapped multiple exons from different transcript isoforms, the mean length of the exons was used. We determined the exon length for each m^6A peak using Bedtools (version v2.25.0) intersectBed and mergeBed. For the RBP internal exon length analysis, a minimum read coverage filter using K562 RNA-seq dataset (Yan et al., 2018) was imposed to filter out intronic binding sites that overlap internal exons from transcript variants not expressed in K562. Statistical significance was determined with the Wilcoxon rank sum test.

RT-qPCR

RT-qPCR was used to assess the relative abundance of mRNA. Total RNA or purified mRNA was reverse transcribed with Primescript Reverse transcriptase (Takara) using poly(dT) primers and random 6mers to obtain complementary DNA. FastStart Essential DNA Green Master Mix (Roche) was used for qPCR reactions, with primers added to 5 μM. qPCR assays were performed on a LightCycler® 96 Instrument (Roche). qPCR was run using the following conditions: 95°C, 10min; (95°C, 20s; 60°C, 20s; 72°C 20s)×40 cycles. Primers used for RT-qPCR can be found in Table 8.

m^{6}A-IP-RT-qPCR

For m^{6}A-IP-RT-qPCR, the m^{6}A-IP procedure used in the m^{6}A-seq was followed until the library preparation step, then input and IP RNA were used for RT-qPCR as described above. m^{6}A enrichment is calculated by taking the IP/input ratio and normalizing to an IP/input ratio for a reference RNA. An m^{6}A-methylated Gaussia luciferase mRNA spike in or endogenous transcripts from the genes *EID1* or *CCDC77* was used for normalization to control for IP efficiency, as indicated in the corresponding figures. If the measured transcript was expressed from a plasmid, prior to m^{6}A-IP, RNA was incubated with 2U of Turbo DNAse in a 50 μl reaction for 37°C for 1 hour and then purified with the RNA Clean and Concentrator-5 kit (Zymo Research).

SELECT

For each sample, 25 ng of total RNA was mixed with 0.8 μl up-probe and down-probe oligonucleotide (1 μM), 1 μl dTTP (100 μM), and 2 μl 10X CutSmart buffer (NEB) supplemented with H_2O to 17 μl total volume. The reaction was incubated at a temperature gradient: 90 °C for

1 min, 80 °C for 1 min, 70 °C for 1 min, 60 °C for 1 min, 50 °C for 1 min, and then 40 °C for 6 min. Subsequently, a 3 µl of enzyme mixture containing 0.5 µl *Bst* 2.0 DNA polymerase (0.02 U/µl) (NEB M0275S), 0.5 µl SplintR ligase (1 U/µl) (NEB M0375S), and 2 µl ATP (5 mM) was added for a final volume of 20 µl. The final reaction mixture was incubated at 40 °C for 20 min then denatured at 80 °C for 20 min. qPCR is used to quantify the abundance of the ligated product. 10 µl 2X FastStart Essential DNA Green Master Mix (Roche) qPCR master mix, 0.8 µl universal SELECT primer (10 µM), 2 µl reaction from previous step, and 7.2 µl H_2O. qPCR quantification of the mRNA levels for the target gene is used to normalize the amount of ligated product for gene expression variation.

Western blot

Cells were lysed in either 2X Laemmli buffer (Biorad) or NP-40 lysis buffer (150 mM sodium chloride, 1.0% NP-40, 50 mM Tris pH 7.5) with 1 × protease inhibitor cocktail (Sigma). Pierce™ BCA Protein Assay Kit (Thermo Fisher) was used to measure protein concentrations. SDS-PAGE was carried out using NuPAGE 4 to 12%, Bis-Tris, 1.5 mm, Mini Protein Gels and buffers (Invitrogen). Proteins were transferred to a .45 µm pore nitrocellulose membrane (Biorad) using a semi-dry transfer apparatus and stained with Ponceau S (Sigma) to check transfer efficiency. Membrane was blocked with PBST (0.1% (v/v) Tween 20 in 1x Phosphate-buffered Saline) with 5% blotting grade blocker (Biorad) for 1 hour at room temperature. Primary antibody with PBST with 2% blocker or BSA and .02% sodium azide was added overnight at 4°C. at dilution listed in Table 8. Membrane was washed three times for 5 min in PBST. Secondary antibody conjugated to HRP added in PBST with 2% blocker for 1 hour at 4°C at dilution listed in Table 8. Washed

3X5min in PBST. Membrane was washed three times for 5 min in PBST. Membrane was visualized using SuperSignal™ West Dura Extended Duration Substrate (Thermo Scientific).

LC-MS/MS quantification of m^6A in RNA

For each sample, 25 ng of RNA was digested with 1 unit of Nuclease P1 (Sigma) in P1 buffer (20 mM NH4OAc, pH 5.5) in a final reaction volume of 20 µL for 2 hours at 42°C. Subsequently, 1µL of FastAP (Thermo Fisher) and 2.5µL of 10x FastAP buffer were added to each sample, and they were incubated at 37°C for 4 hours. Samples were then diluted with 27µL of water and filtered through a 0.2µm PVDF filter (0.2µm pore size, 0.4mm diameter, Millipore). 5µL of each filtered sample was separated by reverse phase ultra-performance liquid chromatography on a C18 column on an Agilent Technologies 1290 Infinity II liquid chromatography system, followed by mass spectrometry on a Sciex Triple Quad 6500 triple-quadrupole mass spectrometer in positive electrospray ionization mode. Nucleosides were quantified using nucleoside-to-base transitions of 282.101>150.100 (m^6A), 267.966>136.000 (A), and 284.004>152.100 (G). Quantification was performed by comparing with the standard curve obtained from pure nucleoside standards running with the same batch of samples. The concentration of m^6A was normalized to the concentration of A or G in each sample to enable comparisons in m^6A levels between samples.

Isolation of EJC-protected RNA footprints

EJC-protected RNA footprints were isolated with a procedure adapted from previous reports (Mabin et al., 2018; Singh et al., 2012). For isolation of EJC-protected RNA footprints via FLAG-Magoh or FLAG-eIF4A3 immunoprecipitation, TREx-HEK293 cells containing a stable copy of FLAG-tagged Magoh or a FLAG-tagged eIF4A3 were used. TREx-HEK293 cells were grown in

four 15-cm plates. Expression of the FLAG-tagged protein was induced with a working concentration of 200 ng/mL tetracycline for ~16 hr. 90 minutes prior to cell harvesting, cycloheximide (CHX) was added to 100 μg/ml. The monolayer was rinsed and harvested in phosphate-buffered saline (PBS) containing 100 μg/ml CHX. The cells were lysed in 4 ml hypotonic lysis buffer (20 mM Tris-HCl pH7.5, 15 mM 11 NaCl, 10 mM EDTA, 0.5% NP-40, 0.1% Triton X-100, 1x Halt Protease Inhibitor (Thermo Fisher), .2 U/μl of SUPERase-In (Thermo Fisher), 100 μg/ml CHX) for 1 hour on ice. The suspension was sonicated (Branson Digital Sonifier-250) at 40% amplitude using a Microtip for a total of 16 seconds (in 2 second bursts with 10 second intervals). NaCl was adjusted to 150 mM and the lysate was cleared by centrifugation at 10,000xg for 10 min at 4°C. The cleared lysate was diluted to 10 ml in the above lysis buffer with final NaCl concentration of 150 mM. The diluted lysate was incubated for 3 hr at 4°C with 1mL of anti-FLAG magnetic beads (50% slurry, Sigma) pre-washed twice with 12 ml IsoWB (Isotonic wash buffer; 20 mM TrisHCl pH7.5, 150 mM NaCl, 0.1% NP-40). The RNA-protein (RNP) complexes captured on beads were sequentially washed four times (4 x 12ml) with ice-cold IsoWB. After the fourth wash, bound RNP complexes were incubated with one bed volume of IsoWB containing 1U/μl of RNase I (Ambion) for 10 min at 37°C shaking at 1200 rpm. RNP complexes were again washed four times with 10 ml IsoWB supplemented with .02 U/μL SUPERase-IN. RNPs were eluted with 40 μl of clear sample buffer (100 mM Tris-HCl pH6.8, 4% SDS, 10 mM EDTA, 100 mM DTT) at 25°C for 5 min, and subsequently at 95°C for 2 min. For isolation of EJC-protected RNA footprints via endogenous eIF4A3 immunoprecipitation, HEK293T cells were harvested and lysed as above, and the cleared lysate was added to 300 μl Protein G Dynabeads pre-coupled to 40 μg of anti-eIF4A3 (Bethyl A302-980A).

Immunoprecipitation, RNase I digestion, and elution were carried out as above. After elution, RNA footprints were isolated by Trizol/chloroform extraction and isopropanol precipitation with GlycoBlue coprecipitant (Invitrogen) and then run on a 6% TBE-Urea gel to assess length distribution. m^6A levels were measured by LC-MS/MS as described above.

In vitro methylation of EJC-protected RNA footprints

For in vitro methylation of EJC-bound RNA eluted under native conditions, the procedure for isolating EJC-protected RNA footprints from TREx-HEK293 cells expressing FLAG-tagged eIF4A3(described above) was followed until the denaturing elution step. RNP complexes were eluted under native conditions by incubating beads with 100 µl of elution buffer consisting of 28 mM Tris-HCl (pH 8.0), 112 mM KCl, 2.1 mM $MgCl_2$, and 1µg/µL FLAG peptide (Sigma) per 250 µl of anti-FLAG 50% magnetic beads slurry used, rotating at 4°C for 2.5 hours. The supernatant containing the eluted RNPs was collected, and DTT, glycerol, SUPERase-IN RNase inhibitor (Invitrogen), SAM (Cayman) and recombinant METTL3-METTL14 (Active Motif) were added up to the following final reaction conditions: 300 nM recombinant METTL3/METTL14 complex, 1.6 mM SAM, 80 mM KCl, 1.5 mM $MgCl_2$, 0.2 U µl−1 SUPERase-IN RNase inhibitor, 10 mM DTT, 4% glycerol and 20 mM Tris-HCl (pH 8.0). For methylation reactions containing RNPs with unmethylated RNA spiked in, in vitro transcribed MPm^6A library (produced as described above) was added to the reaction to a final concentration of 7 ng/ul, approximately 10-fold the yield of eluted EJC-protected RNA footprints. Reactions were incubated at 16°C for 12 hours, then collected in Trizol and purified by Trizol/chloroform extraction and isopropanol precipitation with GlycoBlue coprecipitant. For in vitro methylation of Trizol purified, deproteinized EJC-protected RNA footprints, EJC-protected RNA footprints were first purified

using Trizol/chloroform extraction and isopropanol precipitation with GlycoBlue coprecipitant. The isolated RNA was subjected to the same in vitro methylation conditions as the RNPs eluted under native conditions described above, at 16°C for 12 hours. m^6A levels were measured by LC-MS/MS as described above, subtracting background signal from mock reactions containing all components except for RNA.

Protein co-immunoprecipitation

For METTL3 co-immunoprecipitation, 1e7 HeLa cells were lysed in 500 ul of ice-cold lysis buffer (50 mM Tris-HCl pH 7.4, 100 mM KCl, 5 mM MgCl2, .5% NP-40, 1x Halt Protease Inhibitor). Cells were incubated on ice for 5 min. Lysate was then sonicated in a Bioruptor Pico, with five cycles of 30 second bursts at 30 second intervals. Lysates were spun down, 20,000xg for 10 min at 4°C, and a portion was saved as an input sample. 50 µl M280 sheep anti-rabbit beads (Thermo Fisher) was incubated with 4 µg of anti-METTL3 (Abcam, ab195352), or 4µg of rabbit IgG isotype control(Cell Signaling Technology) rotating end over end for 1 hour at room temperature, and washed 6 times in NT-2 wash buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, .05% NP-40). Lysate was added to prepared beads and incubated rotating end over end at 4°C for 16 hours. Beads were washed 5 times with ice-cold NT-2 wash buffer, and proteins were eluted by addition of 2X Laemmli buffer and incubation at 95°C for 30 min. Input and IP samples were run on an SDS-PAGE gel and visualized by Western blot, as described above. For eIF4A3 co-immunoprecipitation, HEK293T cells were lysed and eIF4A3 was immunoprecipitated as described for above for isolation of EJC-protected footprints, but with a 16-hour incubation of lysate with beads and antibody, rotating end over end at 4°C. Beads were treated with RNase I as above, or left untreated. Proteins were eluted from beads by addition of 2X Laemmli buffer and

incubation at 95°C for 30 minutes. Input and IP samples were run on an SDS-PAGE gel and visualized by Western blot, as described above.

Differential expression analysis

DESeq2 (version 1.18.1) was used to detect differentially expressed genes using a significance threshold of adjusted p-value <.05 (Love et al., 2014). To identify genes that have a significant change in response to EJC KD in WT vs. METTL3$^{mut/-}$ cells, the design formula ~ genotype + treatment + genotype:treatment was used. The P-value and magnitude of the genotype:treatment interaction term for tested genes was reported.

Human tissue m^{6}A analysis

m^{6}A-MeRIP-seq data for human tissues was obtained from Liu et al. 2020 (31). As described in (31), we used read 2 of each fastq file for alignment. Reads were trimmed with TrimGalore (version 0.6.6) with options --length 30 --clip_R1 3 (Martin, 2011). HISAT2 (version 2.1.0) was used to align the trimmed sequence reads to reference genome (hg38) with options -k 1 and --known-splicesite-infile generated by the hisat2_extract_splice_sites.py script with the UCSC hg38 GTF. samtools (version 1.7) was used to generate, sort and index bam files. BigWig files for the forward and reverse strand were created using deeptools (version 3.1.3) bamCoverage with options --binSize 10, --normalizeUsing CPM, --exactScaling, --filterRNAstrand forward/reverse. To assess the methylation status of EJC-suppressed m^{6}A regions in tissues, we took siEIF4A3-1 KD significantly hypermethylated regions from HeLa cells that did not overlap a HeLa siC m^{6}A peak. We filtered out m^{6}A peaks from human tissues that overlapped any portion of siC m^{6}A peaks and filtered out any siEIF4A3-1 KD hypermethylated regions that overlapped any HeLa m^{6}A peak

regions from the Liu et al. 2020 dataset to control for batch effects or protocol-specific differences. We then assessed the methylation status of EJC-suppressed m^{6}A regions in tissues by assessing the overlap of these filtered hypermethylated regions with filtered tissue m^{6}A peaks. To identify peaks that overlapped exon-intron boundaries we filtered for peaks that spanned splice sites and exhibited m^{6}A-IP read coverage > 1 CPM in the intron 20 nt away from splice sites. To obtain the number of m^{6}A-IP reads that mapped to the exon-intron junction, and obtain the m^{6}A enrichment for peaks, we used the summarizeOverlaps function from the GenomicAlignments R package to count reads. We normalized the read numbers to the total number of mapped reads in each sample and set a minimum input expression filter. We identified intron retention events using IRFinder, with default settings, using a cutoff of IR ratio >.1, as was previously described (Middleton et al., 2017). We obtained polyA sites from PolyA.Site.2.0 (Herrmann et al., 2020) and gene models from GENCODE v32 comprehensive transcript set.

m^{6}A-QTL analysis

m^{6}A-QTLs and corresponding m^{6}A peaks were obtained from (Zhang et al., 2020). Reads were aligned as previously described, and BigWig files for the forward and reverse strand were created using deeptools (version 3.1.3) bamCoverage with options --binSize 10, --normalizeUsing CPM, --exactScaling, --filterRNAstrand forward/reverse. To assess the methylation status of EJC-suppressed m^{6}A regions in tissues, we took regions that were significantly hypermethylated upon siEIF4A3-1 and siEIF4A3-2 KD in regions from HeLa and HEK293T cells that did not overlap an siC m^{6}A peak in HeLa and HEK293T cells, respectively. We filtered out m^{6}A-QTL m^{6}A peaks that overlapped with any portion of siC m^{6}A peaks for HeLa and HEK293T cells. We then assessed whether the filtered m^{6}A-QTL m^{6}A peaks overlapped the filtered EJC-suppressed m^{6}A regions.

We obtained polyA sites from PolyA.Site.2.0 (Herrmann et al., 2020) and gene models from GENCODE v32 comprehensive transcript set.

Enrichment tests

For all enrichment tests, Fisher's exact test was used to assess significance of enrichment. To test enrichment of RBPs at MPm^{6}A endogenously methylated vs MPm^{6}A suppressed m^{6}A sites, 120 K562 RBP eCLIP datasets were obtained from ENCODE (Van Nostrand et al., 2020). Peaks with -$\log_{10}$ (pValue) >= 3 and $\log_2$(signalValue) >= 3 in two replicates were used. Benjamin-Hochberg procedure was used to control FDR. Only sites that were not differentially expressed more than 5-fold in K562 (using DESeq2) were included in the analysis to filter possible cell-type specific peaks due to differential gene expression. To test enrichment of FTO and ALKBH5 at MPm^{6}A endogenously methylated vs MPm^{6}A suppressed m^{6}A sites, FTO and ALKBH5 binding sites were obtained from (Baltz et al., 2012; Bartosovic et al., 2017). To test enrichment of eIF4A3-suppressed suppressed splice sites at MPm^{6}A endogenously methylated CDS m^{6}A sites vs MPm^{6}A suppressed CDS m^{6}A sites, and enrichment of eIF4A3-suppressed suppressed splice sites at *EIF4A3* KD hypermethylated regions vs m^{6}A peak regions that are not hypermethylated, eIF4A3-suppressed suppressed splice sites were obtained from (Boehm et al., 2018). 100 nt region proximal to suppressed splice sites was used for enrichment analysis. MPm^{6}A endogenously methylated CDS m^{6}A sites (that remain methylated in the MPm^{6}A reporter) and MPm^{6}A suppressed CDS m^{6}A sites were taken from cellular 3'UTR MPm^{6}A data. *EIF4A3* KD and m^{6}A peak regions that are not hypermethylated were obtained from data from siC and siEIF4A3-1 KD in HeLa cells described in Figure 3. m^{6}A peak regions that are not hypermethylated were obtained by taking m^{6}A peaks that do not overlap *EIF4A3* KD hypermethylated regions. To test enrichment

of m^{6}A-containing transcripts in transcripts with at least one internal exon > 400 nt vs transcripts with all internal exons < 400 nt, constitutive internal exons greater than 400 nt in length were obtained from the HEXEvent database(Busch and Hertel, 2013) with constitLevel > .9. Long internal exons were associated with their respective genes to identify genes that contained at least one internal exon > 400nt. m^{6}A-methylated genes were defined based on HeLa m^{6}A-seq data. Gene expression levels were calculated using Kallisto (version 0.44.0) (Bray et al., 2016) and an expression cutoff at the lowest quartile of expression for all m^{6}A-methylated genes was used to exclude lowly expressed genes that do not have sufficient read coverage to detect m^{6}A peaks from the analysis.

Data and code accessibility

Raw and processed data can be found at NCBI GEO accession GSE162199. Custom scripts available upon request.

RESULTS—EXON ARCHITECTURE CONTROLS mRNA m^{6}A MODIFICATION AND GENE EXPRESSION

Introduction

N^6–methyladenosine (m^{6}A), the most prevalent mRNA modification in mammals, impacts numerous aspects of gene expression and plays essential roles in a wide array of physiological and pathophysiological processes (Frye et al., 2018; Gilbert et al., 2016; Roundtree et al., 2017a). m^{6}A methylation is found at thousands of sites across the transcriptome and regulates many steps in mRNA metabolism, including pre-mRNA processing, nuclear export, mRNA localization, mRNA decay, and translation (He and He, 2021). The METTL3-METTL14 RNA methyltransferase complex catalyzes m^{6}A methylation on mRNA (Bokar et al., 1994; Liu et al., 2014, 2020a; Wang et al., 2016a, 2016b). METTL3-METTL14 installs m^{6}A in a common DRACH (D = A, G or T; R= A or G; H= A, C or U) mRNA sequence motif, but only a fraction of DRACH sequences (~5%) in a subset of cellular transcripts are methylated (He and He, 2021). In addition, m^{6}A exhibits a marked regional bias in its transcriptomic distribution, being strongly enriched in long internal exons and near stop codons (Batista et al., 2014; Dominissini et al., 2012; Ke et al., 2015; Meyer et al., 2015). This characteristic transcriptomic distribution of m^{6}A is observed using a variety of m^{6}A mapping technologies across many different tissues and cell types (Dominissini et al., 2012; Ke et al., 2015; Linder et al., 2015; Liu et al., 2020b; Meyer et al., 2012; Xiao et al., 2019).

Selective deposition of m^{6}A on specific transcripts is central to m^{6}A-mediated regulation of gene expression. It allows for selective regulation of specific transcripts by m^{6}A reader proteins, which preferentially bind m^{6}A-modified mRNAs and exert downstream effects on gene expression (Shi et al., 2019). Additionally, the specific location at which an mRNA is m^{6}A methylated may

also be relevant for gene expression, since the nature of m^{6}A-mediated regulation appears to vary depending on the position of m^{6}A within the transcript (Barbieri et al., 2017; Choi et al., 2016; Mao et al., 2019; Meyer et al., 2015; Slobodin et al., 2017; Wang et al., 2014a, 2015; Zhou et al., 2018).

Despite the central importance of specific m^{6}A deposition in m^{6}A-mediated gene regulation, the mechanistic basis for m^{6}A specificity has remained poorly understood. The METTL3/METTL14 methyltransferase complex accessory subunit VIRMA appears to promote methylation near the stop codon and 3'UTR, but the mechanistic details are not entirely clear (Yue et al., 2018). Notably, most studies that characterize pathways regulating preferential m^{6}A deposition have focused on mechanisms that selectively promote methylation at specific regions. These activating mechanisms involve local recruitment of the methyltransferase by various trans-acting factors, such as transcription factors (TFs), RNA-binding proteins (RBPs), RNA Polymerase II and chromatin modifications, to promote targeted methylation of nearby RNA sequences (Aguilo et al., 2015; Barbieri et al., 2017; Bertero et al., 2018b; Fish et al., 2019; Huang et al., 2019; Slobodin et al., 2017b; Zhang et al., 2020). However, the global enrichment of m^{6}A at specific transcriptomic features such as long internal exons and near stop codons, as well as selective deposition of m^{6}A on a large portion of methylated transcripts, are not explained by known pathways, suggesting that major regulatory mechanisms that control m^{6}A deposition and confer m^{6}A epitranscriptome specificity are still unknown.

In this study, we develop a Massively Parallel Assay for m^{6}A (MPm^{6}A) and apply it to systematically uncover determinants of m^{6}A on an epitranscriptome-wide scale. We discover, contrary to our initial expectations, the existence of prevalent regulatory mechanisms that restrict m^{6}A methylation to specific transcript regions through targeted suppression of m^{6}A in

unmethylated regions. We find that pre-mRNA splicing selectively suppresses m^{6}A deposition in average-length exons, but not in longer exons. We identify Exon Junction Complexes (EJC), deposited upstream of exon boundaries by spliceosomes, as m^{6}A suppressors that mediate this splicing and exon length-dependent suppression and control multiple key characteristics of global m^{6}A specificity. EJC depletion results in pervasive aberrant methylation of mRNAs, leading to m^{6}A-mediated gene expression dysregulation. A subset of EJC-suppressed m^{6}A sites can escape silencing in specific contexts due to methylation of transcript isoforms with longer exons, which contributes to methylation variation across tissues and inter-individual variation in m^{6}A associated with complex traits. Finally, we find that EJCs collaborate with the peripheral EJC factor RNPS1 to package and protect long stretches of proximal RNA from m^{6}A deposition and that suppressed methylation sites co-localize with EJC/RNPS1-suppressed splice sites in average-length exons (Blazquez et al., 2018; Boehm et al., 2018). This suggests that exon architecture broadly determines local RNA accessibility to mRNA regulatory machineries due to packaging of exon junction-proximal RNA by EJCs.

Massively Parallel Assay for m^{6}A

The extent to which global m^{6}A specificity is controlled by the intrinsic preference of the methyltransferase for specific RNA sequences ("intrinsic determinants") versus the regulation of methyltransferase activity by trans-acting factors ("extrinsic determinants") has important implications for m^{6}A regulation but is poorly understood. One recent report proposed that intrinsic determinants primarily control m^{6}A methylation, and that m^{6}A levels are largely "hard coded" in cis by the local sequence surrounding the m^{6}A site and account for m^{6}A specificity (Garcia-Campos et al., 2019). We addressed this question by asking: is the local sequence surrounding an

m^{6}A methylated site, when uncoupled from its endogenous context, sufficient to specify methylation at that site? And conversely, is the local sequence surrounding an unmethylated DRACH site, when uncoupled from its endogenous context, sufficient to prevent methylation at that site?

To assess this systematically on an epitranscriptome-wide scale, we developed an approach that enables high-throughput assessment of the m^{6}A methylation status of thousands of designed sequences, which we term Massively Parallel assay for m^{6}A (MPm^{6}A) (Figure 4A). In the MPm^{6}A workflow, thousands of endogenously methylated m^{6}A sites, or endogenously unmethylated DRACH sites, and their local flanking sequences are synthesized and then cloned into a plasmid-based transgene. The sequences are expressed and then m^{6}A methylated through transfection into cells or through *in vitro* transcription and *in vitro* m^{6}A methylation. The methylation status of each individual sequence is assessed by its enrichment following m^{6}A-immunoprecipitation (IP) of mRNA, determined by massively parallel sequencing. If local sequence is sufficient to intrinsically encode methylation levels, one would expect that methylated and unmethylated sequences should each retain their endogenous methylation states when expressed within the "artificial" context of the transgene. Conversely, if sequences deviate from their endogenous methylation states when expressed in an artificial context, this would indicate the activity of extrinsic determinants.

We selected 6,897 HeLa m^{6}A sites and 3,058 unmethylated DRACH sites to assay (see Methods). For each site, we synthesized a 102-nucleotide sequence containing the methylated/unmethylated site in the center, flanked by 50/51 nt of adjacent endogenous sequence context. For each sequence, we designed a corresponding negative control sequence in which all DRACH motifs are mutated to prevent methylation. All sequences were associated with three

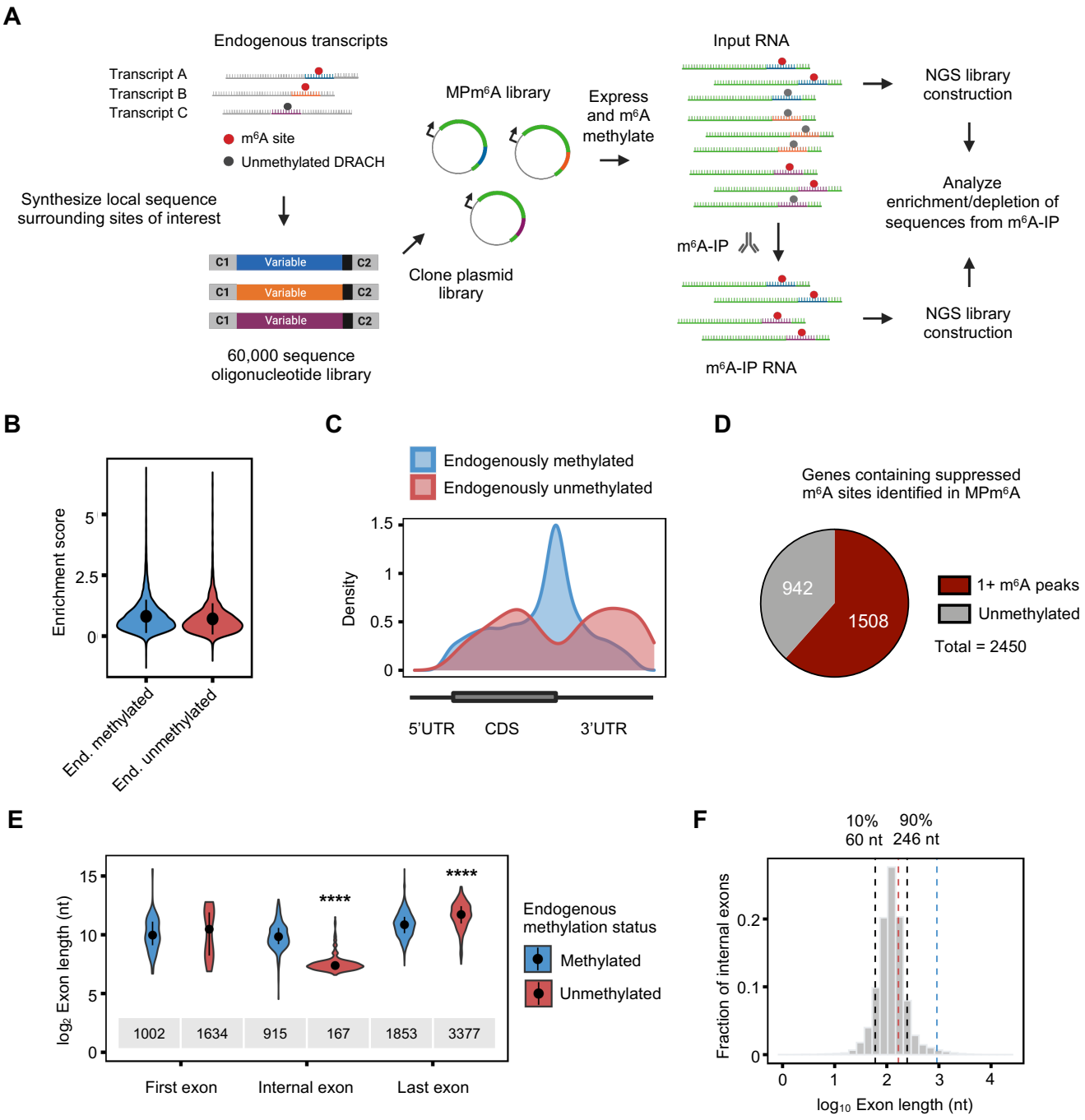

Figure 4. MPm^{6}A reveals widespread suppression of thousands of m^{6}A sites in unmethylated transcriptome regions.

Figure 4, continued
A, Schematic of the MPm^{6}A workflow. **B,** MPm^{6}A enrichment scores (experimental IP/input – negative control IP/input) for endogenously methylated and endogenously unmethylated sequences, mean ± SD, four biological replicates. **C,** Metagene plots of the endogenous locations of endogenously methylated and unmethylated MPm^{6}A sequences that are significantly methylated (P<.05) in MPm^{6}A. **D,** Number of genes containing suppressed m^{6}A sites identified in MPm^{6}A that exhibit one or more m^{6}A peaks on other regions of the endogenous transcript, and number that do not exhibit m^{6}A on any part of the transcript. **E,** Exon lengths of endogenously methylated and endogenously unmethylated sequences that are significantly methylated in MPm^{6}A, median and IQR, Wilcoxon rank sum test, ****P<2.2e-16. **F,** Distribution of internal exon lengths in the human genome. Black lines indicate 10th percentile (60 nt) and 90th percentile (246 nt), blue and red lines indicate median internal exon length for MPm^{6}A endogenously methylated (915 nt) and unmethylated sequences (167 nt), respectively**.**

unique barcodes, for a total of 59,730 unique sequences. These sequences were synthesized, cloned into the 3'UTR of a GFP transgene, and transfected into HeLa cells for expression and methylation. 52,866 sequences exceeded the minimum read cutoff and were used for further analysis. We observed high correlations in m^6A enrichment ratios between four biological replicates and observed a strong correlation in m^6A level measurements between MPm^6A and conventional m^6A-IP qPCR (Figure 5A and 5B). As expected, negative control sequences were depleted overall following m^6A-IP (Figure 5C).

Widespread mRNA m^6A suppression controls m^6A epitranscriptome specificity

When we compared the methylation level of the endogenously methylated sequences to their negative control sequences, we found that 92.8% of the sequences exhibited significant methylation in the artificial context of this reporter assay (Figures 4B and 5D). This indicates that most endogenously methylated sites do not strictly require their larger surrounding native context for methylation. We note that the modification fraction may differ between the sequences in the context of the reporter and in their endogenous context, so this does not exclude the possibility that surrounding context affects modification stoichiometry.

Next, we examined the methylation levels of the endogenously unmethylated sequences. Strikingly, 90.2% of these sequences also exhibited significant methylation (Figures 4B and 5D). The enrichment scores of the endogenously unmethylated sequence group were similar to the endogenously methylated group, despite their diverging endogenous methylation states (Figure 4B). We observed similar results when the sequences were *in vitro* transcribed and *in vitro* methylated with recombinant METTL3-METTL14 (Figure 6). Thus, thousands of endogenously

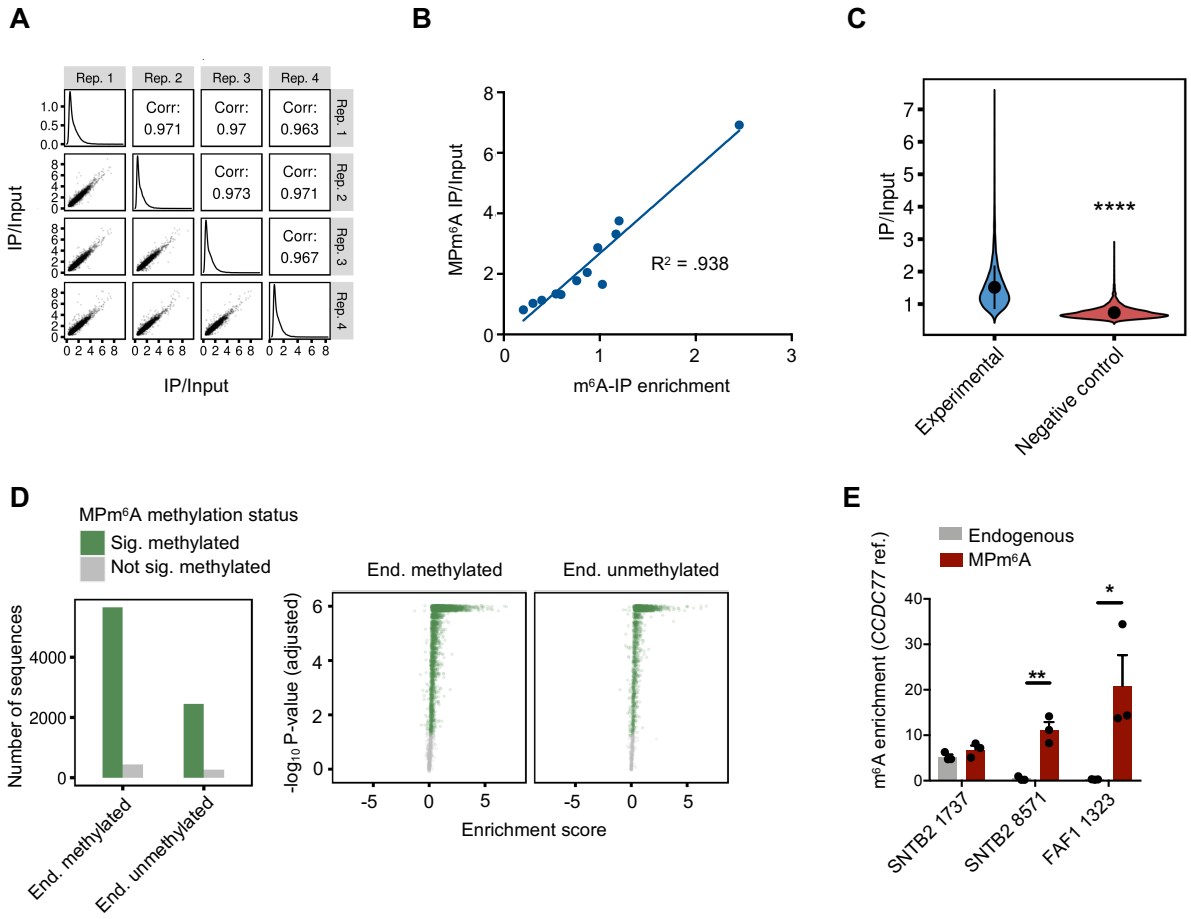

Figure 5. MPm^{6}A reveals widespread suppression of thousands of m^{6}A sites in unmethylated transcriptome regions, continued.

A, Scatterplots of 3'UTR MPm^{6}A (sequences cloned into 3'UTR of reporter plasmid) IP/Input for four biological replicates (bottom left). Histograms of IP/Input for individual replicates (middle). Pearson correlation for biological replicates (top right). **B,** Scatterplot of MPm^{6}A (3'UTR) IP/Input and m^{6}A-IP-qPCR normalized enrichment values for 12 MPm^{6}A sequences. **C,** MPm^{6}A (3'UTR) IP/Input for experimental and negative control mutant sequences, mean ± SD, Wilcoxon rank-sum test, ****P<2.2e-16. **D,** Number of endogenously methylated and endogenously unmethylated sites that are significantly methylated (P <.05, Wilcoxon rank sum test with BH procedure to control FDR) or not significantly methylated (P >.05) in MPm^{6}A (3'UTR), four biological replicates (left). MPm^{6}A (3'UTR) enrichment score and p-values for endogenously methylated and endogenously unmethylated sequences. Significance threshold defined at P<.05, Wilcoxon rank-sum test with BH procedure to control FDR (right). **E,** m^{6}A methylation at *SNTB2* 1737, *SNTB2* 8571, and *FAF1* 1323 sequences in endogenous mRNA or in RNA expressed from the MPm^{6}A reporter (cloned into 3'UTR), assessed by m^{6}A-IP-qPCR, mean ± SEM, two-tailed T-test, *P<0.05; **P<0.01, m^{6}A enrichment is normalized to *CCDC77* mRNA, three biological replicates. Numbers following the gene names indicate the position of the m^{6}A site or unmethylated DRACH site relative to the transcription start site.

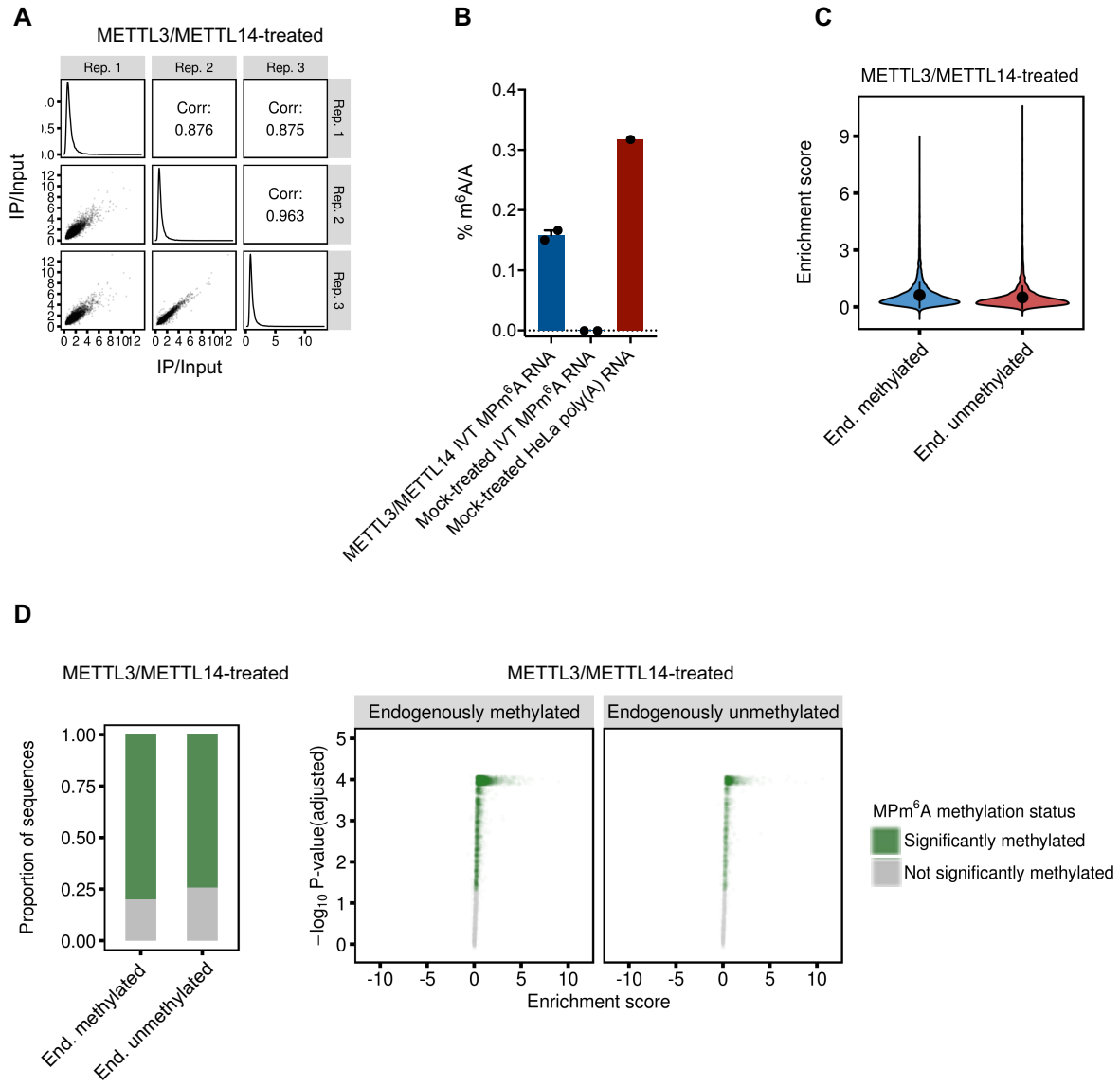

Figure 6. *in vitro* MPm^{6}A reveals that local sequence context is not sufficient to specify endogenous methylation status.

A, Scatterplots of recombinant METTL3/METTL14-treated *in vitro* methylated MPm^{6}A (3'UTR) IP/Input for three replicates (bottom left). Histograms of IP/Input for individual replicates (middle). Pearson correlation between replicates (top right). **B,** LC-MS/MS measurement of m^{6}A levels of *in vitro* transcribed and recombinant METTL3/METTL14-treated *in vitro* methylated MPm^{6}A mRNA, mock-treated (all *in vitro* methylation reaction components, but omitting recombinant METTL3/METTL14) *in vitro* transcribed MPm^{6}A mRNA, and mock-treated cellular polyadenylated RNA. **C,** Enrichment scores for endogenously methylated and endogenously unmethylated sequences for recombinant METTL3/METTL14-treated *in vitro* MPm^{6}A (3'UTR), mean ±SD. **D,** Proportion of sequences that are significantly methylated (P <.05) or not significantly methylated (P >.05) in recombinant METTL3/METTL14-treated *in vitro* MPm^{6}A

Figure 6, continued
(3'UTR). Wilcoxon rank sum test with BH procedure to control FDR, three replicates (left). Enrichment scores and p-values for endogenously methylated and unmethylated sequences in recombinant METTL3/METTL14-treated *in vitro* MPm^{6}A (3'UTR). Significance threshold defined at P<.05, Wilcoxon rank-sum test with BH procedure to control FDR, three replicates (right).

unmethylated DRACH sites become methylated when they are uncoupled from their endogenous contexts and expressed in an artificial reporter context. We term these identified sites "suppressed m^{6}A sites". We validated these results by using m^{6}A-IP-qPCR to measure m^{6}A enrichment at an endogenously methylated site (SNTB2 1737) and two endogenously unmethylated sites (SNTB2 8571, FAF1 1323) in endogenous mRNA and in mRNA expressed from the MPm^{6}A plasmid (Figure 5E).

We also performed variations of MPm^{6}A in which we cloned the 59,730 sequences into the 5'UTR and CDS of the transgene and expressed them in HeLa cells. Consistent with our previous results, the proportion of significantly methylated sequences in the endogenously methylated and endogenously unmethylated groups were similar (Figures 7A-D), confirming the presence of thousands of suppressed m^{6}A sites in the endogenously unmethylated mRNA. m^{6}A enrichment was significantly lower for many sequences when placed in the CDS or 5'UTR versus in the 3'UTR, reaffirming the existence of m^{6}A regulatory mechanisms not encoded within local sequence features (Figure 7E). This may indicate that 5' regions of mRNAs are generally less conducive than 3' regions for m^{6}A methylation. Collectively, these results show that intrinsic local sequence features surrounding DRACH sites are insufficient to encode m^{6}A epitranscriptome specificity, and reveal, for the first time, the existence of thousands of suppressed m^{6}A sites that are silenced by unknown mechanisms.

We next investigated the characteristics of these suppressed m^{6}A sites. We visualized the distribution of the original positions of the assayed MPm^{6}A sequences within their endogenous transcripts. Interestingly, suppressed m^{6}A sites are enriched in the CDS and in the 3'UTR distal to the stop codon and are depleted near the stop codon, forming an inverse distribution to the endogenous m^{6}A sites (Figure 4C). Thus, suppressed m^{6}A sites are globally enriched in

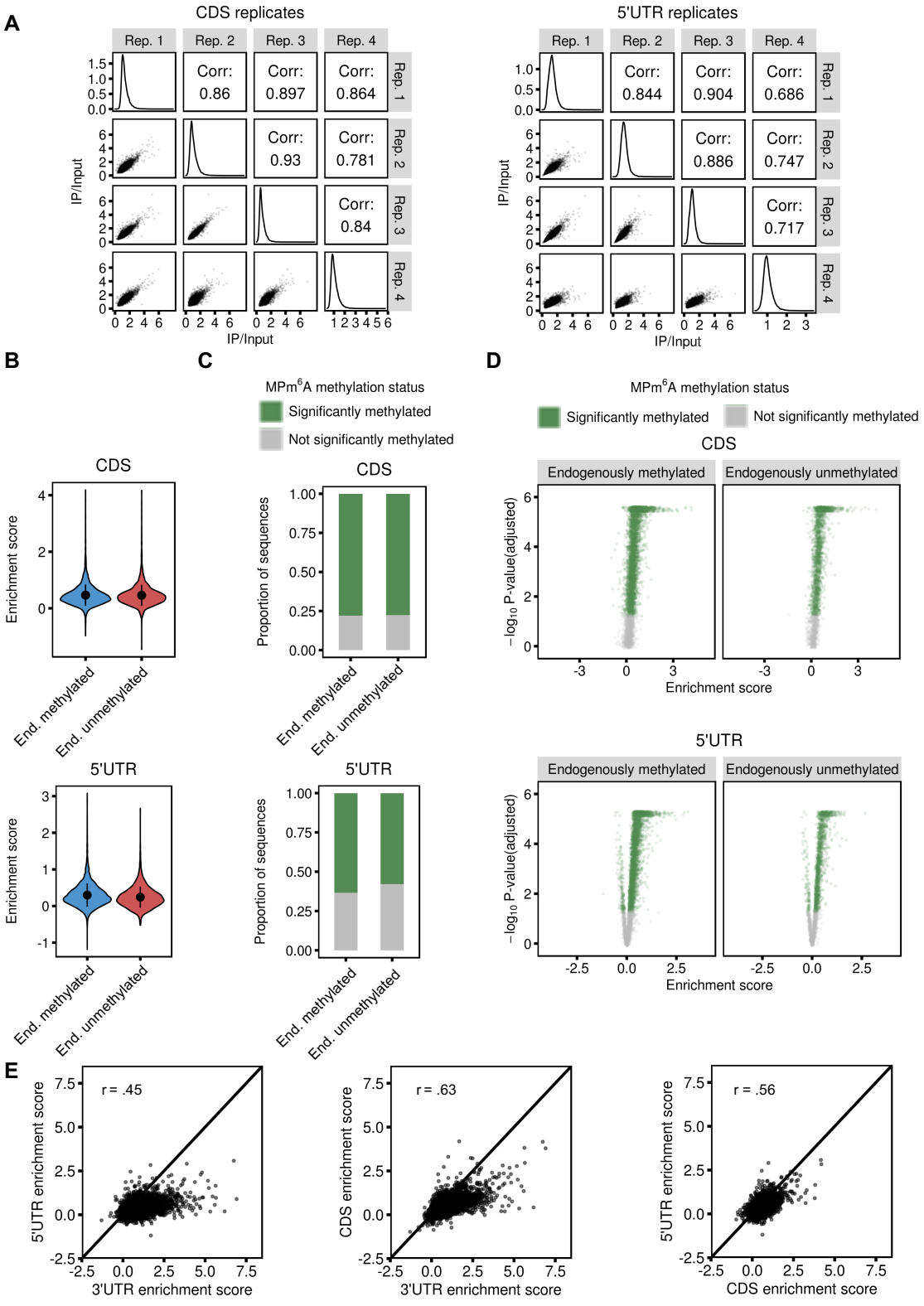

Figure 7. CDS MPm^{6}A and 5'UTR MPm^{6}A reveal that local sequence context is not sufficient to specify endogenous methylation status.

Figure 7, continued
A, Scatterplots of CDS MPm^{6}A IP/Input (left) and 5'UTR MPm^{6}A IP/Input (right) for four biological replicates, (bottom left). Histograms of IP/Input for individual replicates (middle). Pearson correlation for biological replicates (top right). **B,** CDS MPm^{6}A enrichment scores (top) and 5'UTR MPm^{6}A enrichment scores (bottom) for endogenously methylated and endogenously unmethylated sequences, mean ±SD. **C,** Fractions of endogenously methylated and endogenously unmethylated sequences that are significantly methylated ($P < .05$, Wilcoxon rank sum test with BH procedure to control FDR) or not significantly methylated ($P > .05$) for CDS MPm^{6}A (top) and 5'UTR MPm^{6}A (bottom), four biological replicates. **D,** CDS MPm^{6}A (top) and 5'UTR MPm^{6}A (bottom) enrichment scores and p-values for endogenously methylated and endogenously unmethylated sequences. Significance threshold defined at $P<.05$, Wilcoxon rank-sum test with BH procedure to control FDR. **E,** Scatterplots of 5'UTR and 3'UTR MPm^{6}A enrichment scores (left), CDS and 3'UTR MPm^{6}A enrichment scores (middle), 5'UTR and CDS MPm^{6}A enrichment scores (right), for endogenously methylated sequences. A $y = x$ line is plotted for reference.

transcriptome regions in which m^{6}A is depleted. We found that 942 genes containing suppressed m^{6}A sites do not contain any endogenous m^{6}A methylation on their transcripts (Figure 4D). Further, suppressed m^{6}A sites in internal exons reside within much shorter exons (median = 167 nt) than endogenous m^{6}A sites (median = 915 nt) (Figure 4E). This suggests that enrichment of m^{6}A in long internal exons may derive from suppression of m^{6}A sites in shorter internal exons, which comprise the vast majority of internal exons (90% of internal exons are < 246 nt) (Figure 4F). These results collectively suggested that widespread suppression of these sites contributes significantly to m^{6}A epitranscriptome specificity.

The extensive m^{6}A suppression we observed was unexpected initially, given that nearly all previously described pathways of m^{6}A regulation are activating mechanisms (Barbieri et al., 2017; Bertero et al., 2018b; Fish et al., 2019; Huang et al., 2019; Slobodin et al., 2017b). No known mechanisms appear to account for this widespread targeted suppression of m^{6}A deposition. These unknown suppressive mechanisms appear to involve suppression of m^{6}A deposition by the m^{6}A writer complex rather than active demethylation of these sites by m^{6}A erasers, since binding sites for RBM15, a METTL3/METTL14 methyltransferase complex accessory subunit, are highly enriched near the endogenous m^{6}A sites compared to the suppressed m^{6}A sites (OR = 33.7) (Figure 8A). 153 RBM15 binding sites overlap endogenous m^{6}A sites while only 2 overlap suppressed m^{6}A sites. In contrast, FTO and ALKBH5 binding sites were not significantly enriched near suppressed sites (FTO OR = 0.8, ALKBH5 OR = 1.3), and they exhibited little binding near suppressed sites overall (Figure 8B) (Baltz et al., 2012; Bartosovic et al., 2017). Therefore, our MPm^{6}A assay suggests the existence of unknown m^{6}A "suppressor" proteins that govern global m^{6}A specificity by suppressing m^{6}A deposition.

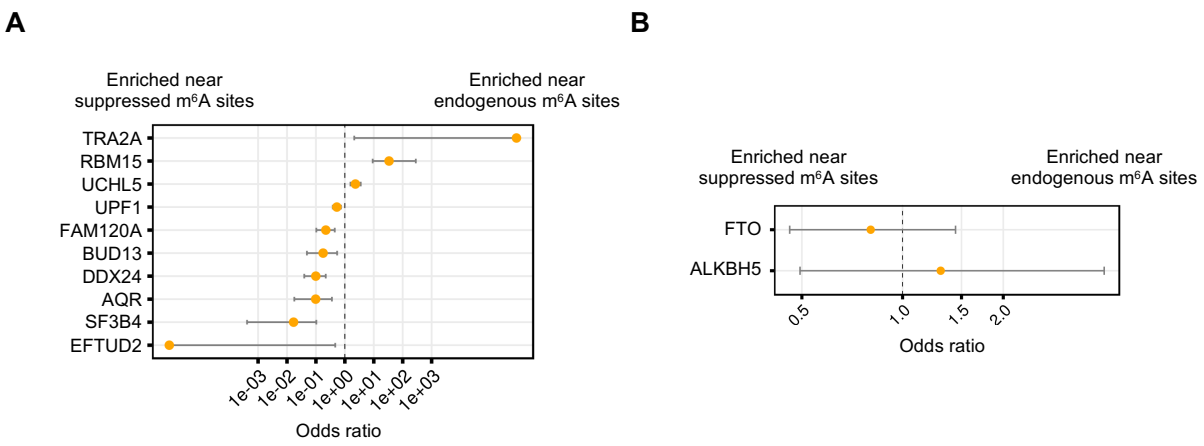

Figure 8. Spliceosome components are enriched near suppressed m^{6}A sites.

A, RBPs with statistically significant enrichment or depletion (P<.05, Fisher's exact test with BH procedure to control FDR) of binding sites at endogenously methylated or endogenously unmethylated MPm^{6}A sequences that are significantly methylated in MPm^{6}A. Dot and bar represent odds ratio and 95% confidence interval. **B,** Enrichment of FTO and ALKBH5 binding sites at endogenously methylated or endogenously unmethylated MPm^{6}A sequences that are significantly methylated in MPm^{6}A. Dot and bar represent odds ratio and 95% confidence interval.

Pre-mRNA splicing strongly suppresses m^{6}A methylation in proximity to splice sites

To identify factors that mediate the suppression of these m^{6}A sites, we examined the relative enrichment of binding sites for 120 RBPs at endogenously methylated vs. suppressed m^{6}A sites (Van Nostrand et al., 2020). We identified three RBPs that were significantly enriched at endogenously methylated m^{6}A sites, including RBM15 as previously mentioned, and seven RBPs that were significantly enriched at suppressed m^{6}A sites (Figure 8A). Interestingly, several components of the spliceosome (*BUD13, SF3B4, EFTUD2*) were significantly enriched at suppressed sites. This result is concordant with the fact that the CDS suppressed m^{6}A sites primarily reside within average-length internal exons, and therefore are in close proximity to both upstream and downstream splice sites (Figure 4E). Thus, we hypothesized that the splicing of average-length internal exons may suppress m^{6}A methylation within these exons. To test this, we cloned a suppressed m^{6}A site from an average-length internal exon in the *CRY1* gene (Figure 9A) into a rabbit beta-globin minigene reporter (BG), as well as a version with the introns removed (BG Δi1,i2). First, we cloned the suppressed *CRY1* site and 50/51nt of flanking exonic sequence into the internal exon, or in the last exon of these constructs. Notably, the spliced construct strongly suppressed methylation of the m^{6}A site when placed within the internal exon, but not within the last exon (Figure 9B). Removal of either the upstream or downstream intron (BG CRY1 Δi1 102, BG CRY1 Δi2 102) both resulted in partial loss of suppression, indicating that splicing of both introns is required for complete suppression (Figure 10A). Deletion of all splice sites abolished suppression, indicating that the act of splicing, rather than specific intronic sequences, is required for suppression (Figure 9C). Cloning in 912 nt of the *CRY1* exonic sequence surrounding the suppressed site into the internal exon (BG CRY1 912), forming a long internal exon, resulted in a total loss of suppression (Figure 10A). Based on these results, we hypothesized that the

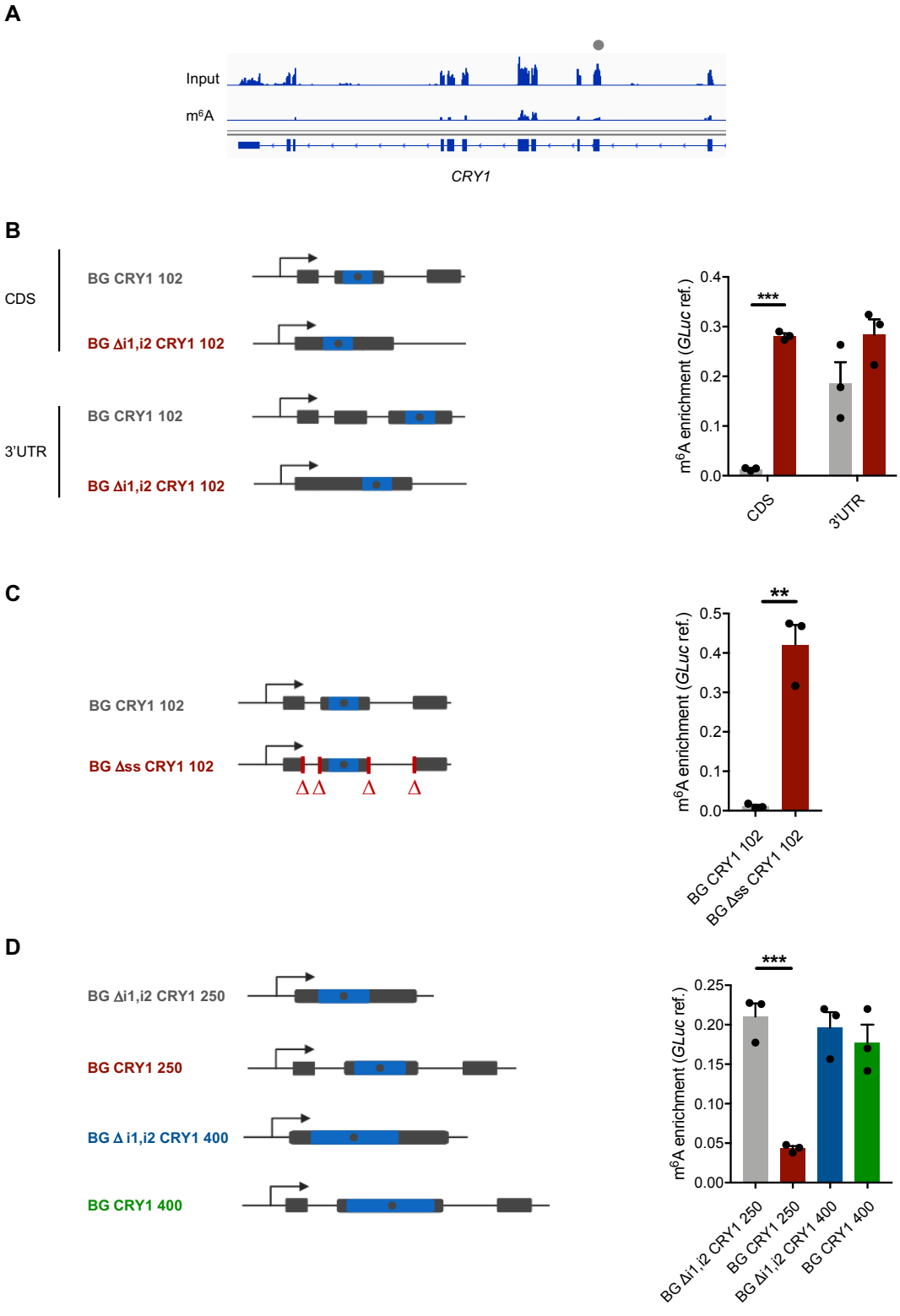

Figure 9. Pre-mRNA splicing suppresses m^{6}A methylation in an exon length-dependent manner.

Figure 9, continued
A, Input and m^{6}A-IP read coverage at the *CRY1* gene in HeLa cells, gray dot marks location of the suppressed m^{6}A site cloned into BG *CRY1* constructs. **B,** m^{6}A methylation of specified BG *CRY1* constructs, with the CRY1 suppressed m^{6}A site cloned into the CDS or 3'UTR of BG construct containing introns or with introns removed. Schematic of constructs (left), gray dot represents position of the *CRY1* suppressed m^{6}A site. Blue region indicates sequence derived from *CRY1*, gray region indicates sequence derived from rabbit beta-globin. Number following CRY1 refers to nucleotides of exonic sequence surrounding the *CRY1* suppressed m^{6}A site in the endogenous *CRY1* mRNA that is cloned into BG constructs. m^{6}A enrichment at the *CRY1* suppressed m^{6}A site within mRNA expressed from the constructs (right). **C,** m^{6}A methylation of specified BG *CRY1* constructs, with 102 nt of exonic sequence surrounding the *CRY1* suppressed m^{6}A site cloned into the internal exon of BG CRY1, or with 102 nt of exonic sequence surrounding the *CRY1* suppressed m^{6}A site cloned into the internal exon of BG Δss CRY1, in which all the splice sites are deleted. Schematic of constructs (left), m^{6}A enrichment at the *CRY1* suppressed m^{6}A site (right). **D,** m^{6}A methylation of specified BG *CRY1* constructs with the presence of introns and length of internal exon varied. Schematic of constructs (left), m^{6}A enrichment at the *CRY1* suppressed m^{6}A site (right). For B to D, RNA is sheared to ~150 nt fragments and primers amplify 62 nt fragment containing the *CRY1* suppressed m^{6}A site. mean ± SEM , two-tailed T-test, **, P<.01, ***, P<.001, three biological replicates.

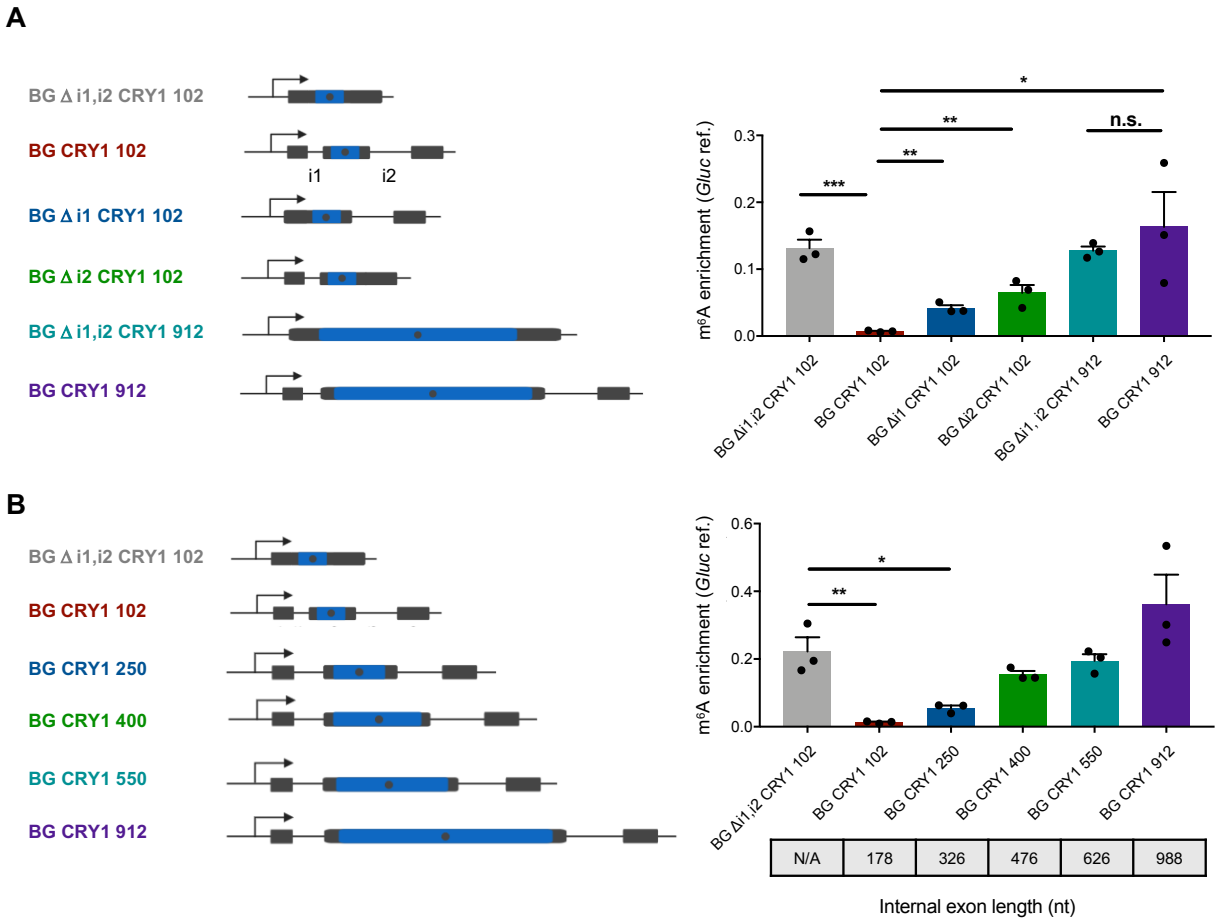

Figure 10. Pre-mRNA splicing suppresses m^{6}A methylation in an exon length-dependent manner, continued.

A, m^{6}A methylation of specified BG CRY1 constructs. Schematic of constructs (left), gray dot represents the position of the *CRY1* suppressed m^{6}A site. Blue region indicates sequence derived from the *CRY1* endogenous sequence surrounding the suppressed m^{6}A site, gray regions indicate sequence derived from rabbit beta-globin (BG). Number following CRY1 refers to the number of nucleotides of exonic sequence surrounding the *CRY1* suppressed m^{6}A site in the *CRY1* endogenous mRNA that is cloned into the BG construct. Δ denotes deletion of the specified intron(s). BG CRY1 102, BG Δ i1,i2 CRY1 102, BG Δi1 CRY1 102, and BG Δi2 CRY1 102 all contain 102 nt of endogenous *CRY1* exonic sequence surrounding the suppressed m^{6}A site within their internal exons. BG CRY1 102 contains both introns, BG Δ i1,i2 CRY1 102 lacks both introns, and BG Δ i1 CRY1 102 and BG Δ i2 CRY1 102 lack the first and second introns, respectively. BG CRY1 912 and BG Δ i1,i2 CRY1 912 contain 912 nt of endogenous *CRY1* exonic sequence surrounding the suppressed m^{6}A site within their internal exons. BG CRY1 912 contains both introns while BG Δ i1,i2 CRY1 912 lacks both introns. m^{6}A enrichment at the *CRY1* suppressed m^{6}A site (right), mean ± SEM, two-tailed T-test, *P<0.05; **P<0.01, ***P<.001, three biological replicates. **B,** m^{6}A methylation of specified BG CRY1 constructs.

Figure 10, continued
Schematic of constructs (left), BG CRY1 102, BG Δ i1,i2 CRY1 102 and BG CRY1 912 as described in a. BG CRY1 250, BG CRY1 400, BG CRY1 550 contain 250, 400, and 550 nt of endogenous *CRY1* exonic sequence surrounding the suppressed m^6A site within their internal exons, respectively, and contain both introns. m^6A enrichment at the *CRY1* suppressed m^6A site (right), mean ± SEM, two-tailed T-test, *P<0.05; **P<0.01, three biological replicates. For A and B, RNA is sheared to ~150 nt fragments and primers amplify a 62 nt fragment containing the *CRY1* suppressed m^6A site. m^6A enrichment is calculated as IP/input normalized to m^6A^+ Gaussia luciferase RNA spike-in IP/input.

suppression is dependent on the proximity of the m^{6}A site, located within the center of the exon, to splice sites. Expanding the length of the BG CRY1 102 internal exon by cloning in larger amounts of flanking sequence resulted in a progressive loss of suppression, with a >= 476 nt internal exon unable to suppress m^{6}A (Figures 10B and 9D). These results reveal, for the first time, a causal role for splicing events in regulating m^{6}A. Splicing can suppress m^{6}A in an exon length-dependent manner, strongly suppressing m^{6}A sites within internal exons near average length, but not in longer internal exons or terminal exons. Because the vast majority of human internal exons are shorter than 476 nt, this mechanism could suppress thousands of CDS suppressed m^{6}A sites that reside in proximity to splice sites.

Exon junction complexes control m^{6}A epitranscriptome specificity

To further elucidate this m^{6}A suppression pathway, we next searched for specific cellular factors that mediate strong splicing-dependent suppression of m^{6}A specifically within average-length exons. One candidate factor we considered was the exon junction complex (EJC). The EJC is deposited by spliceosomes onto mRNA ~24 nt upstream of exon-exon junctions and plays many roles in gene expression regulation (Boehm and Gehring, 2016; Hir et al., 2016). Notably, two recent studies report that EJCs efficiently block splicing at proximal aberrant splice sites (Blazquez et al., 2018; Boehm et al., 2018). Additionally, EJCs package and compact mRNA and can protect long stretches of proximal RNA from nuclease accessibility in vitro, and also block 5' to 3' exonuclease degradation in vivo (Lee et al., 2020; Singh et al., 2012). We reasoned that suppressed m^{6}A sites within average-length internal exons are within relatively close proximity to both an upstream and downstream EJC. Conversely, m^{6}A sites within long internal exons and near stop codons (which generally reside in long last exons) can be hundreds of nucleotides removed from

the nearest EJC and escape suppression. We therefore hypothesized that EJCs could mediate the splice site-proximal suppression of m^{6}A we observed.

We knocked down (KD) the core EJC factor eIF4A3 in HeLa cells and assessed the effect on m^{6}A deposition transcriptome-wide using m^{6}A-MeRIP-seq. Notably, 24,350 regions were significantly hypermethylated upon *EIF4A3* KD, while 3,140 regions were hypomethylated (Figure 11A). 9,457 of these hypermethylated regions exhibited a greater than 8-fold increase in m^{6}A enrichment compared to the non-targeting siRNA control. To assess whether these m^{6}A changes are mediated by the EJC, we knocked down RBM8A, another core EJC factor (Tange et al., 2005). We observed similar, though relatively milder, transcriptome-wide m^{6}A changes, with 14,034 significantly hypermethylated regions observed, of which 8,012 overlapped with hypermethylated regions observed in *EIF4A3* KD (Figure 11A, 12A and 12B). The relatively milder m^{6}A changes upon *RBM8A* KD may result from relatively lower KD efficiency (Table 2) or may indicate a stronger requirement of eIF4A3 for suppression. Concordant with these transcriptome-wide m^{6}A changes, using LC-MS/MS, we found that *EIF4A3* KD increased global levels of m^{6}A in polyadenylated RNA by two-fold, while *RBM8A* KD resulted in a ~25% increase (Figure 12C).

22,927 hypermethylated regions from *EIF4A3* KD and 11,557 hypermethylated regions from *RBM8A* KD do not overlap with m^{6}A peaks identified in the non-targeting siRNA control condition, suggesting that these regions contain newly methylated suppressed m^{6}A sites (Figure 11A). These newly methylated sites were observed on both unmethylated regions of endogenously methylated transcripts (e.g. *SRSF6, DDX3X*) as well as endogenously unmethylated transcripts

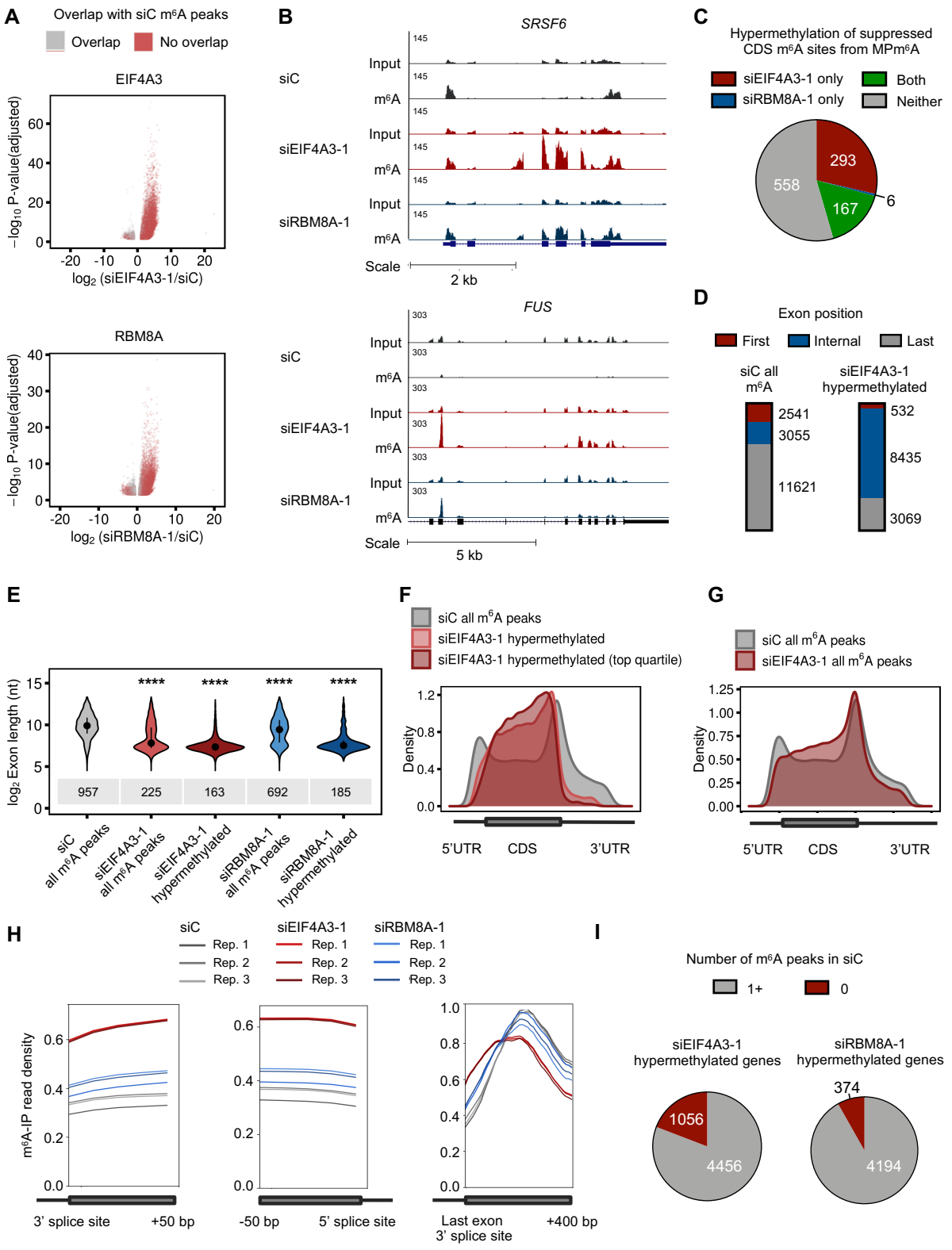

Figure 11. EJCs protect exon junction-proximal RNA in average-length exons within coding sequences from m^{6}A methylation.

Figure 11, continued

A, Differentially methylated regions upon *EIF4A3* knockdown (KD) in HeLa cells (FDR<.1, $|\log_2 FC|>1$), three biological replicates (top). Differentially methylated regions upon *RBM8A* knockdown (KD) in HeLa cells (FDR<.1, $|\log_2 FC|>1$), three biological replicates (bottom). Gray dots indicate differentially methylated regions that overlap m^6A peaks in the non-targeting control KD cells (siC), red dots indicate differentially methylated regions that do not overlap m^6A peaks in the control KD. **B,** Input and m^6A-IP read coverage at *SRSF6* and *FUS* upon non-targeting control KD, *EIF4A3* KD and *RBM8A* KD in HeLa cells. **C,** Overlap of suppressed CDS m^6A sites identified in MPm^6A with hypermethylated regions upon *EIF4A3* and/or *RBM8A* KD. **D,** Number of all m^6A peaks in control KD and number of *EIF4A3* KD hypermethylated regions that unambiguously reside within first, internal or last exons. **E,** Exon lengths of all m^6A peaks residing within internal exons in control KD, *EIF4A3* KD and *RBM8A* KD, and exon lengths of m^6A hypermethylated regions within internal exons upon *EIF4A3* and *RBM8A* KD in HeLa cells. Dot and bar represent median and interquartile range, Wilcoxon rank sum test of indicated group vs. siC all m^6A peaks, ****P<2.2e-16. **F,** Metagene plots of all m^6A peaks upon control KD and m^6A hypermethylated regions upon *EIF4A3* KD in HeLa cells. **G,** Metagene plots of all m^6A peaks upon control KD and all m^6A peaks upon *EIF4A3* KD in HeLa cells. **H,** Density of m^6A-IP reads 50 nt upstream and downstream 5' and 3' splice sites, respectively, and 400 nt downstream last exon 3' splice sites upon *EIF4A3* KD and *RBM8A* KD. **I,** Number of hypermethylated genes upon *EIF4A3* KD or *RBM8A* KD with at least one m^6A peak in the control KD, and number of hypermethylated genes with no m^6A peaks in the control KD. Hypermethylated genes are defined as genes whose transcripts contain a new m^6A peak upon *EIF4A3* or *RBM8A* KD that is not present in the control KD.

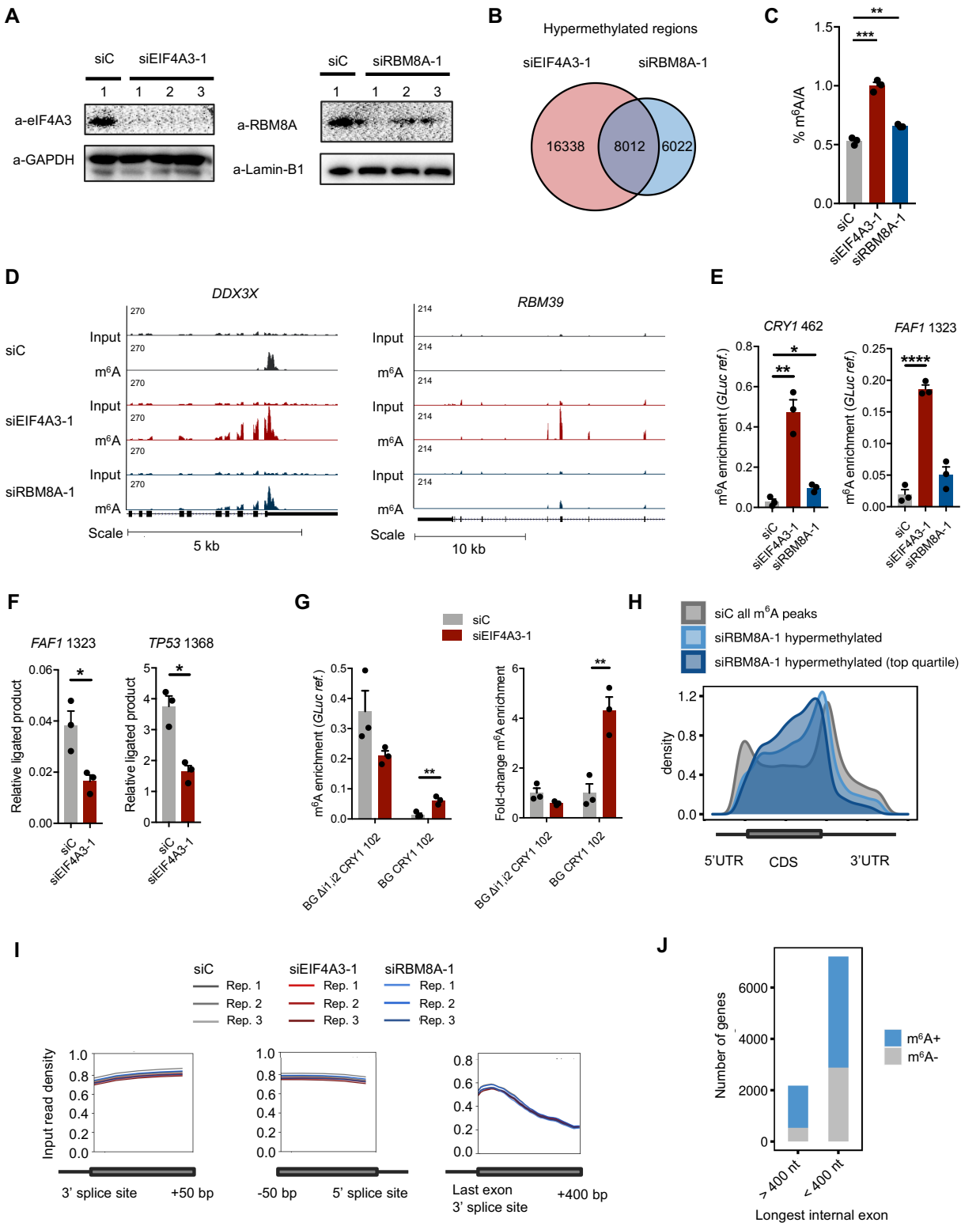

Figure 12. EJCs protect exon junction-proximal RNA in average-length exons within coding sequences from m^{6}A methylation, continued.

Figure 12, continued
A, Knockdown efficiency of siEIF4A3-1 and siRBM8A-1 in HeLa cells. **B,** Overlap of hypermethylated regions upon *EIF4A3* and *RBM8A* KD in HeLa cells. **C,** LC-MS/MS measurement of m^6A levels of polyadenylated RNA from *EIF4A3* and *RBM8A* KD HeLa cells, mean ± SEM, two-tailed T-test, ***P=.0002, **P=.0027. **D,** Input and m^6A-IP read coverage at *DDX3X* and *RBM39* upon non-targeting control KD, *EIF4A3* KD and *RBM8A* KD in HeLa cells. **E,** m^6A methylation at *CRY1* 462 and *FAF1* 1323 suppressed m^6A sites upon *EIF4A3* and *RBM8A* KD in HeLa cells assessed by m^6A-IP-qPCR. 462 and 1323 denote the position of the site relative to the transcription start site, mean ± SEM, two-tailed T-test, ****P<.0001, **P=.0019, *P=.0163. **F,** m^6A methylation at *FAF1* 1323 and *TP53* 1368 suppressed m^6A sites upon *EIF4A3* KD in HeLa cells assessed by SELECT, mean ± SEM, two-tailed T-test, *P<.05. Relative ligated product levels are inversely proportional to m^6A levels; a decrease in relative ligated product indicates an increase in m^6A. **G,** m^6A methylation enrichment at the *CRY1* suppressed m^6A site in specified BG *CRY1* constructs upon EIF4A3 KD. Results depicted as normalized m^6A enrichment (left) and fold-change compared to control KD (right). Number following *CRY1* refers to nucleotides surrounding the *CRY1* suppressed m^6A site in the endogenous *CRY1* mRNA that is cloned into the BG construct. Δi1,i2 denotes deletion of both introns. RNA is sheared to ~150 nt fragments and primers amplify 62 nt fragment containing *CRY1* suppressed m^6A site. m^6A level is normalized to m^6A^+ *in vitro* transcribed Gaussia luciferase RNA spike in, mean ±SEM, two-tailed T-test, **, P<0.01, three biological replicates**. H,** Metagenes of m^6A hypermethylated regions upon *RBM8A* KD in HeLa cells. **I,** Density of input reads 50 nt upstream and downstream 5' and 3' splice sites, respectively, and 400 nt downstream last exon 3' splice sites upon *EIF4A3* KD and *RBM8A* KD. **J,** Transcript methylation status of genes in HeLa cells, grouped by presence of an internal exon > 400 nt. m^6A+ indicates that transcripts of these genes contain at least one m^6A peak, m^6A- indicates that transcripts of these genes do not contain an m^6A peak.

(e.g. *FUS, RBM39*), and reside predominantly within average-length internal exons within coding sequences (Figure 11B and 12D). Indeed, out of 1,024 CDS sequences identified by MPm^{6}A to contain suppressed m^{6}A sites, 466 become methylated upon *EIF4A3* and/or *RBM8A* KD, including the *CRY1* suppressed site (Fig. 11C). This observation indicates that EJCs silence many of the suppressed m^{6}A sites identified by MPm^{6}A. We confirmed methylation of suppressed m^{6}A sites by m^{6}A-IP-qPCR and using SELECT, an antibody-independent m^{6}A measurement method (Figure 12E and 12F) (Xiao et al., 2018). Further, we found that *EIF4A3* KD substantially alleviates the previously observed m^{6}A suppression within the internal exon of BG CRY1 102, further indicating that the EJC mediates the suppressive effect of splicing on m^{6}A deposition (Figure 12G).

Consistent with our model, newly methylated and hypermethylated regions resulting from *EIF4A3* and *RBM8A* KD are highly enriched in average-length internal exons (medians = 163 nt, 185 nt) within CDSs (Figure 11E-G and 12H). Indeed, *EIF4A3* and *RBM8A* KD both resulted in transcriptome-wide increases in m^{6}A enrichment in exon junction-proximal regions (Figure 11H and 12I). *EIF4A3* KD disrupts m^{6}A epitranscriptome specificity globally, resulting in substantial loss of enrichment of m^{6}A peaks in long internal exons (siC median = 957 nt, siEIF4A3 median = 225 nt) and increased density of m^{6}A in the CDS relative to the stop codon (Figure 11E-G). It was previously reported that the peak of m^{6}A density near stop codons on metagene plots can be more precisely visualized as an increase in enrichment 150 nt past the start of last exons, which we also observe (Figure 11H) (Ke et al., 2015). *EIF4A3* KD results in a global increase in m^{6}A enrichment <150 nt past the start of last exons (Figure 11H and 12I). This suggests that EJC suppression of methylation proximal to last exon-exon junctions is responsible for the notable increase in m^{6}A enrichment >150 nt past the start of last exons, near stop codons.

Most transcripts exhibiting hypermethylation upon *EIF4A3* KD have one or more m^{6}A peaks under the non-targeting siRNA control conditions. However, over a thousand transcripts that ordinarily lack m^{6}A peaks also gain aberrant m^{6}A methylation upon *EIF4A3* KD, revealing a major role for EJCs in suppressing m^{6}A deposition on the subset of transcripts that ordinarily are not subject to m^{6}A regulation (Figure 11I). This is consistent with the fact that genes without long internal exons, which constitute the majority of genes, are less likely to have m^{6}A-methylated transcripts than those with long internal exons (Figure 12J). We depleted eIF4A3 with a different siRNA in HeLa cells as well as with both *EIF4A3* siRNAs in HEK293T cells and observed similar transcriptome-wide m^{6}A changes in each case (Figures 13 and 14). Altogether, these data indicate that spliceosomes widely suppress m^{6}A methylation via deposition of EJCs that protect proximal RNA from methylation. This protection underlies multiple characteristics of global m^{6}A epitranscriptome specificity, including enrichment of m^{6}A in long internal exons, depletion of m^{6}A in CDSs and enrichment of m^{6}A in last exons near stop codons, and methylation selectivity for transcripts possessing long internal exons.

mRNA expressed from unspliced expression constructs is hypermethylated

Expression constructs that lack endogenous splice sites are widely used as biological research tools and in gene therapy. The pervasive m^{6}A suppression by EJCs suggests that many mRNAs expressed from unspliced transgenes may be aberrantly hypermethylated compared to their endogenous transcripts. Indeed, we found that mRNAs from thousands of distinct plasmid transgenes are aberrantly hypermethylated due to lack of physiologic m^{6}A suppression by the EJC (Figures 4 and 10). Interestingly, we noted that *IDS*, which encodes iduronate 2-sulfatase, an enzyme involved in the lysozomal degradation of dermatan sulfate and heparan sulfate, and *SMN1,*

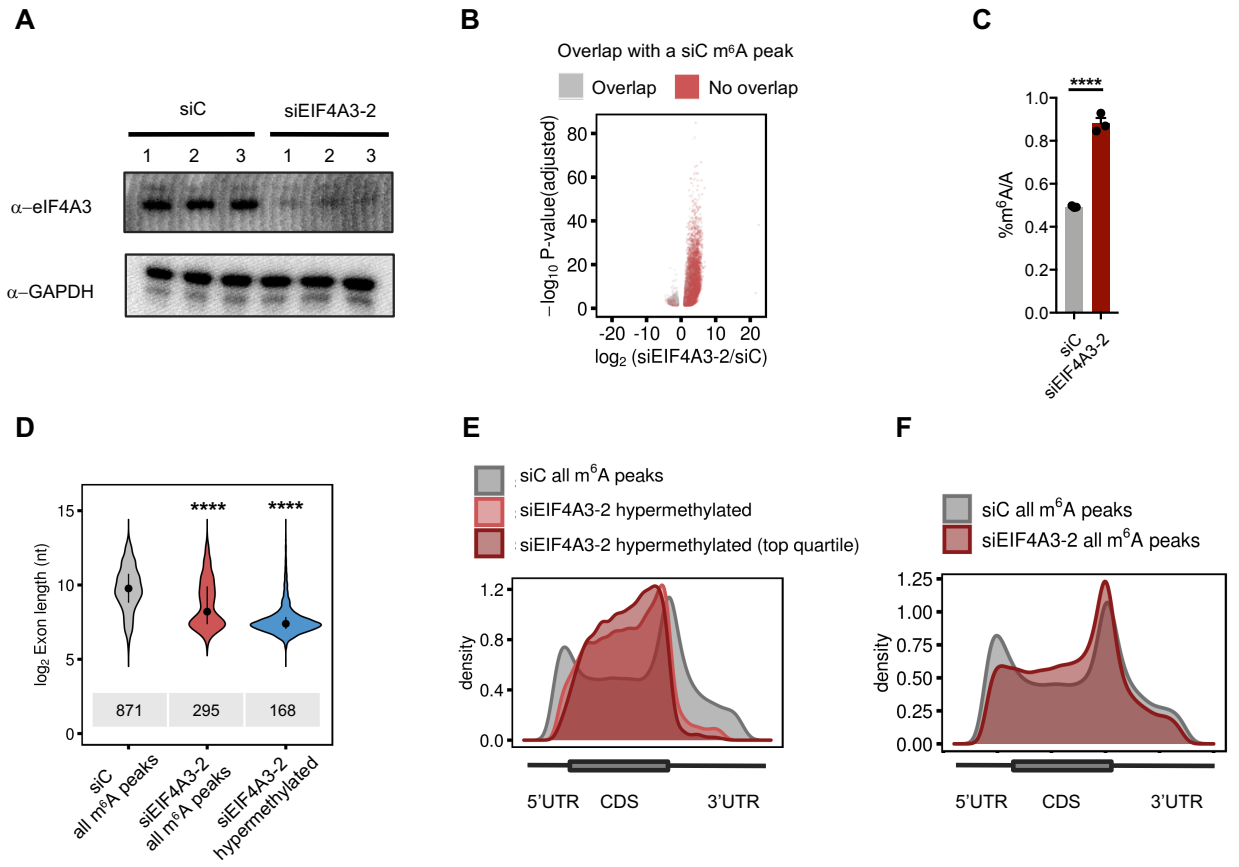

Figure 13. Knockdown of eIF4A3 with an alternative siRNA confirms eIF4A3-mediated suppression of exon junction-proximal m^{6}A methylation.

A, Knockdown efficiency of siEIF4A3-2 in HeLa cells. **B,** Differentially methylated regions upon siEIF4A3-2 knockdown (KD) in HeLa cells (FDR<.1, |log$_2$FC|>1), three biological replicates. Gray dots indicate differentially methylated regions that overlap m^{6}A peaks in the non-targeting control knockdown cells, red dots indicate differentially methylated regions that do not overlap m^{6}A peaks in the control. **C,** LC-MS/MS of m^{6}A of polyadenylated RNA from siEIF4A3-2 HeLa cells, mean ± SEM, two-tailed T-test, ****P<.0001. **D,** Exon lengths of all m^{6}A peaks within internal exons in control and upon KD with siEIF4A3-2, and exon lengths of m^{6}A hypermethylated regions within internal exons upon KD with siEIF4A3-2. Dot and bar represent median and interquartile range, Wilcoxon rank sum test, ****P<2.2e-16. **E,** Metagenes of m^{6}A hypermethylated regions upon KD with siEIF4A3-2 in HeLa cells. **F,** Metagenes of all m^{6}A peaks in control and siEIF4A3-2 KD.

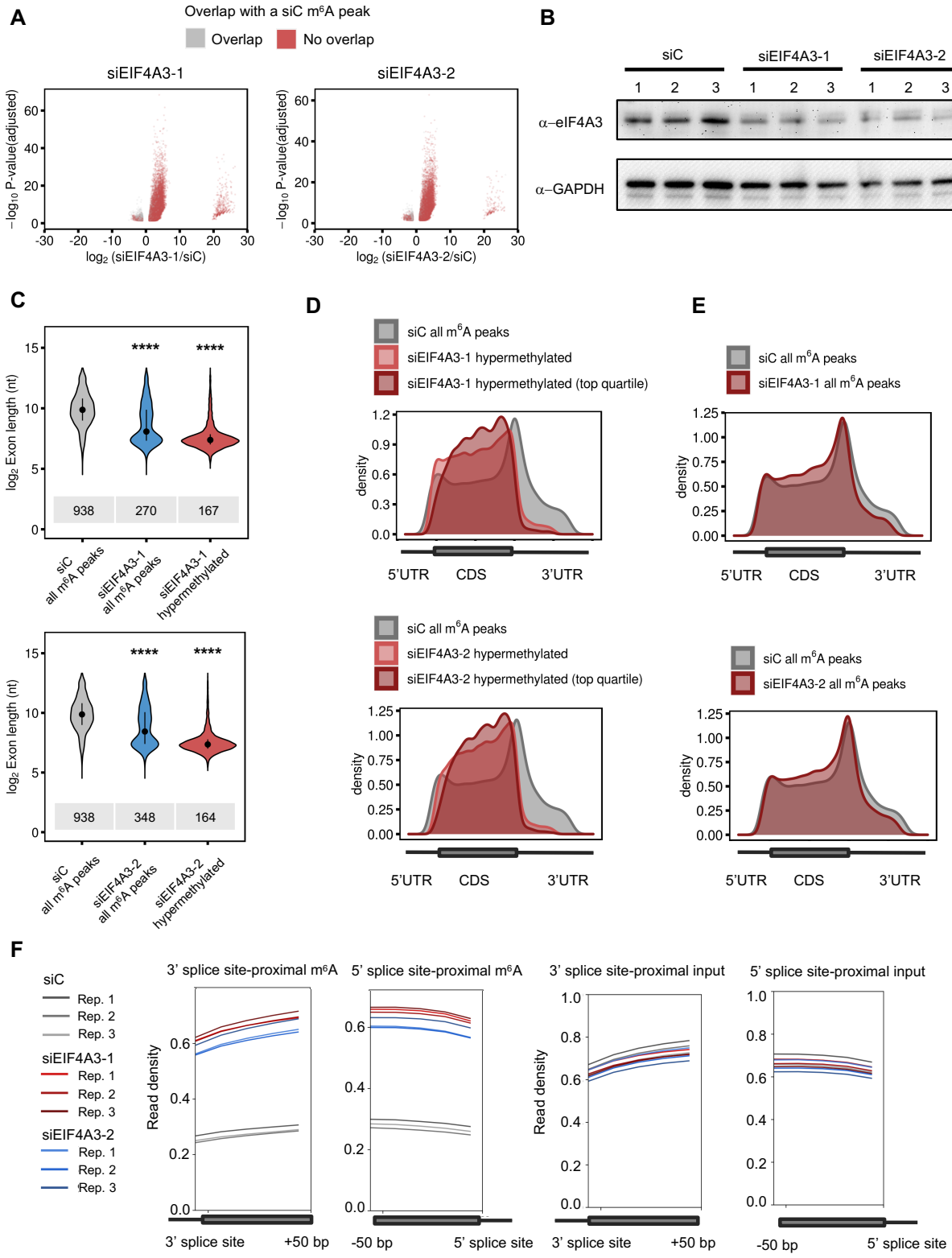

Figure 14. eIF4A3 globally suppresses exon junction-proximal m^{6}A methylation in HEK293T cells.

Figure 14, continued
A, Differentially methylated regions upon *EIF4A3* KD in HEK293T (FDR<.1, $|\log_2 FC|>1$) with two different siRNAs, three biological replicates. Gray dots indicate differentially methylated regions that overlap m^6A peaks in the non-targeting control knockdown cells, red dots indicate differentially methylated regions that do not overlap m^6A peaks in the control. **B,** Knockdown efficiency of siEIF4A3-1 and siEIF4A3-2 in HEK293T cells. **C,** Exon lengths of all m^6A peaks within internal exons in control and upon *EIF4A3* KD, and exon lengths of m^6A hypermethylated regions within internal exons upon *EIF4A3* KD in HEK293T. Dot and bar represent median and interquartile range, Wilcoxon rank sum test of specified group vs. siC all m^6A peaks, ****P<2.2e-16. **D,** Metagenes of all m^6A peaks in control and m^6A hypermethylated regions upon *EIF4A3* KD in HEK293T cells. **E,** Metagenes of all m^6A peaks in control and of all m^6A peaks upon *EIF4A3* KD in HEK293T cells. **F,** Density of m^6A-IP and input reads near 5' and 3' splice sites upon *EIF4A3* KD in HEK293T cells.

which encodes SMN, a protein involved in small nuclear ribonucleoprotein assembly, contain EJC-suppressed m^{6}A sites in their mRNAs, respectively (Figure 15A). Mutations in *IDS* and *SMN1* cause Mucopolysaccharidosis type II (MPSII, also known as Hunter Syndrome, an X-linked lysosomal storage disease) and Spinal Muscular Atrophy (an autosomal recessive motor neuron disease), respectively, and cause significant morbidity and mortality in affected patients (Mercuri et al., 2020; Scarpa, 1993). A gene therapy that delivers a functional copy of *IDS* is currently in clinical trials for MPSII (Ledford, 2018), and a gene therapy that delivers a functional copy of *SMN1* has recently been approved for the treatment of Spinal Muscular Atrophy (AveXis, Inc., 2020). We hypothesized that mRNAs expressed from *IDS* and *SMN1* cDNA expression constructs may be aberrantly methylated due to lack of EJC suppression. Indeed, we observed that mRNAs expressed from cDNAs are significantly hypermethylated relative to the endogenous mRNAs (Figure 15B). Thus mRNA expressed from transgenes that lack endogenous mRNA exon architecture can be hypermethylated due to lack of EJC protection.

EJCs modulate gene expression by suppressing m^{6}A

The most significantly enriched gene ontology terms for transcripts containing EJC-suppressed m^{6}A sites relate to cell division, splicing, and chromatin modification, which are cellular processes that are disrupted upon loss of the EJC (Akhtar et al., 2019; Silver et al., 2010; Wang et al., 2014c) (Figure 16A). This suggests that aberrant m^{6}A hypermethylation upon EJC KD may alter expression of these transcripts and contribute to defects in these pathways. To assess whether m^{6}A mediates some of the gene expression changes observed upon EJC depletion, we knocked down eIF4A3 in both wild-type and *METTL3* heterozygous knockout (*METTL3*$^{-/mut}$)

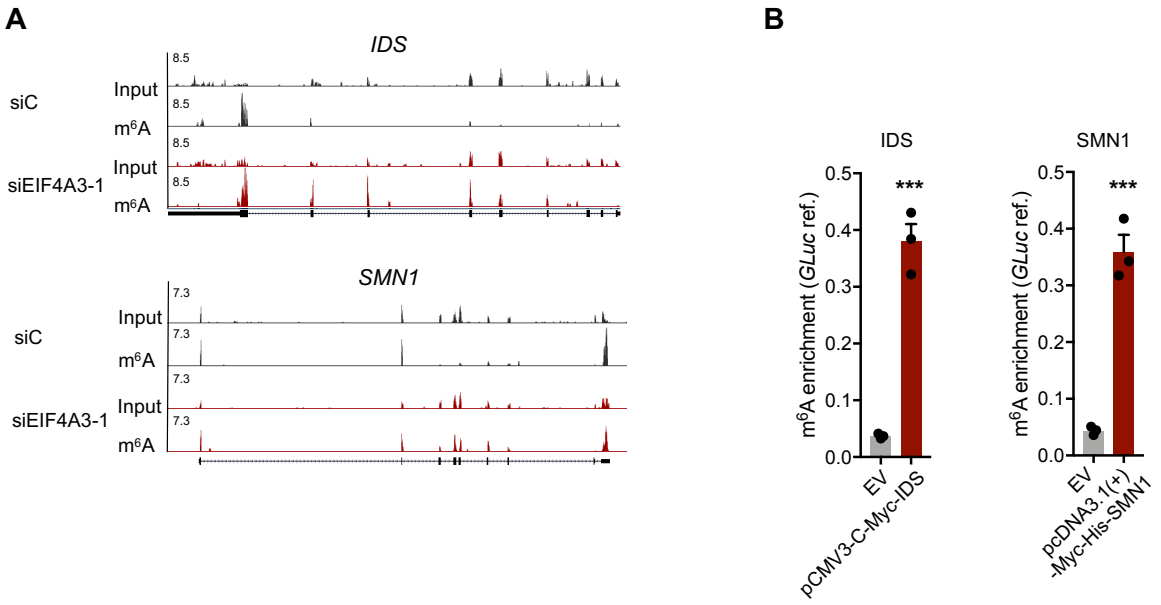

Figure 15. *IDS* and *SMN1* mRNA expressed from unspliced cDNA expression constructs are hypermethylated compared to endogenous mRNA.

A, Input and m^{6}A-IP read coverage at *IDS* and *SMN1* upon non-targeting control KD and *EIF4A3* KD in HeLa cells. **B,** m^{6}A enrichment on *IDS* and *SMN1* mRNA expressed from cDNA expression plasmids, or on endogenous mRNA (corresponding empty vector (EV) plasmid control transfection). m^{6}A enrichment is calculated as IP/Input, normalized by IP/Input of a m^{6}A+ Gaussia luciferase spike in.

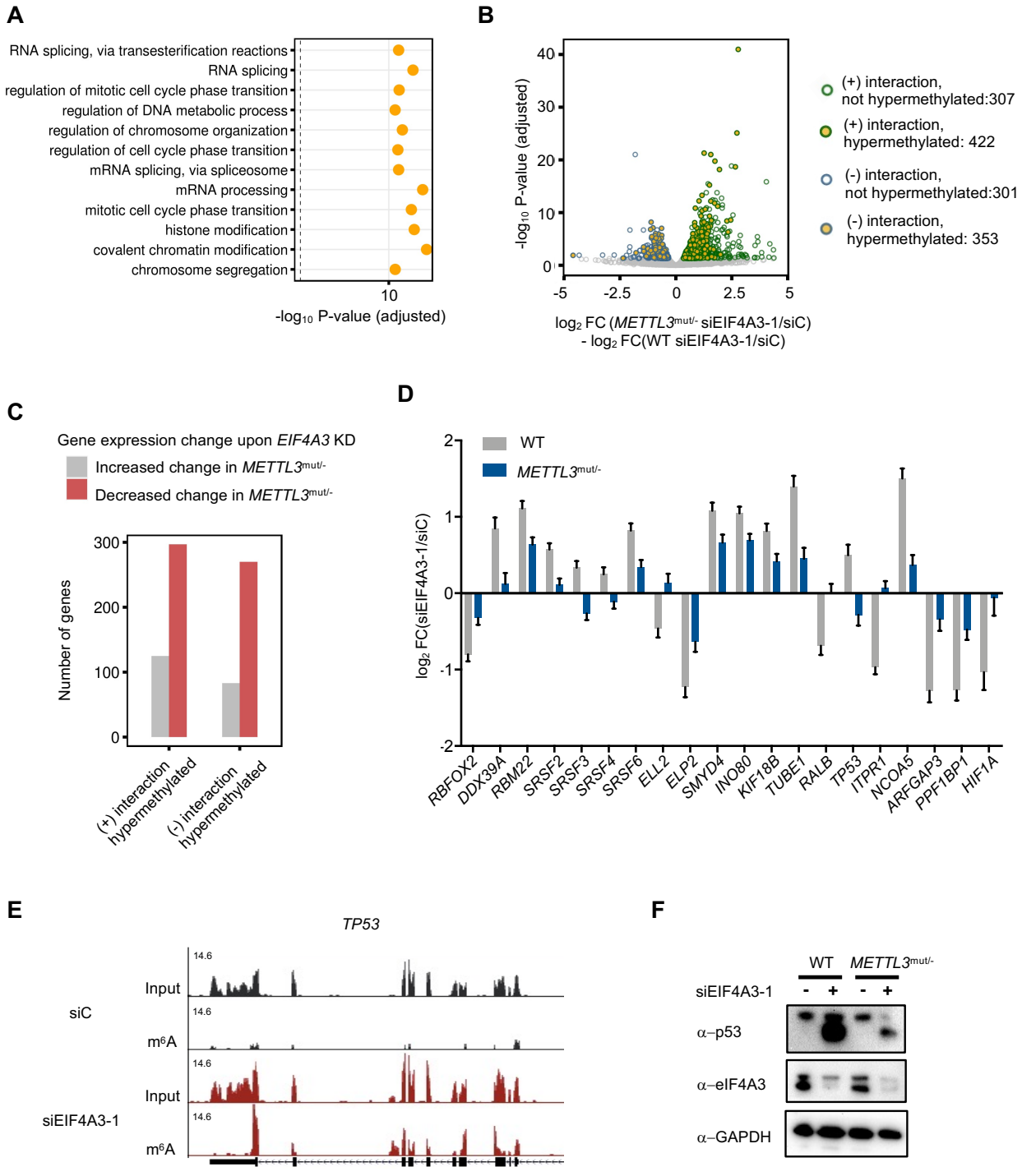

Figure 16. Differential m^{6}A methylation of transcripts upon exon junction complex depletion leads to gene expression dysregulation.

Figure 16, continued

A, Gene ontology term analysis for hypermethylated transcripts upon *EIF4A3* KD, 12 statistically significant terms with lowest P-values depicted. **B,** Differential expression analysis upon *EIF4A3 KD* in WT and *METTL3*$^{mut/-}$ HeLa cells. (+) and (-) interaction denotes genes for which the expression change upon *EIF4A3* KD is significantly (P<.05) more positive or negative, respectively, in *METTL3*$^{mut/-}$ cells compared to change in gene expression upon *EIF4A3* KD in WT cells. Hypermethylated genes are indicated by gold-filled circles. Three biological replicates. **C,** Numbers of genes with hypermethylated transcripts with significant (P<.05) (+) and (-) interactions that exhibit increased or decreased changes in *METTL3*$^{mut/-}$ as compared to WT. For (+) interactions, "increased change" denotes genes for which expression change is positive upon *EIF4A3* KD in WT, and is even more positive in *METTL3*$^{mut/-}$, while "decreased change" denotes genes for which expression change is negative upon *EIF4A3* KD in WT, and less negative or positive in *METTL3*$^{mut/-}$. For (-) interactions, "increased change" denotes genes for which expression change is negative upon *EIF4A3* KD in WT, and is even more negative in *METTL3*$^{mut/-}$, while "decreased change" denotes genes for which expression change is positive upon *EIF4A3* KD in WT, and is less positive or negative upon *EIF4A3* KD in WT *METTL3*$^{mut/-}$. **D,** Selected genes that are hypermethylated upon *EIF4A3* KD for which gene expression is significantly different (adjusted P-value < .05, Wald test) upon *EIF4A3* KD in WT vs. *METTL3*$^{mut/-}$ HeLa cells, mean ± standard error. **E,** Input and m^{6}A-IP read coverage of *TP53* upon *EIF4A3* KD in HeLa cells. **F,** Western blot of p53 levels upon *EIF4A3* KD in wild-type and *METTL3*$^{mut/-}$ cells, representative image of three biological replicates.

HeLa cells (Yue et al., 2018) that exhibit reduced mRNA m^{6}A levels, and performed RNA-seq. We found that for 1,383 genes, the effect of *EIF4A3* KD on gene expression substantially differs in *METTL3*$^{-/mut}$ versus WT cells (Figure 16B). Notably, 775 of these genes have transcripts that are newly methylated or hypermethylated upon *EIF4A3* KD. 353 of these hypermethylated transcripts exhibit a more negative expression change in *METTL3*$^{-/mut}$ vs. WT ((-) interaction) and 422hypermethylated transcripts exhibit a more positive gene expression change ((+) interaction). For most (567) of these hypermethylated transcripts, METTL3 depletion markedly reduced the changes normally observed upon *EIF4A3* KD, suggesting that m^{6}A contributes to the regulatory effects of the EJC on these genes (Fig. 16C). For 270 hypermethylated transcripts, METTL3 depletion notably diminished the upregulation normally observed upon *EIF4A3* KD, while for 297 hypermethylated transcripts METTL3 depletion markedly diminished the downregulation normally observed upon *EIF4A3* KD, suggesting that m^{6}A hypermethylation resulting from EJC depletion exerts positive or negative effects on gene expression in a transcript-dependent manner. METTL3 depletion significantly reduced the changes typically observed upon *EIF4A3* KD for hypermethylated mRNAs encoding splicing regulators (*RBFOX2*, *DDX39A*, *RBM22*, *SRSF2*, *SRSF3*, *SRSF4*, *SRSF6*), chromatin modifiers and transcription elongation regulators (*ELL2*, *ELP2*, *SMYD4*, *INO80*), and proteins involved in cell division (*KIF18B*, *TUBE1*, *RALB*) (Figure 16D). Thus, m^{6}A hypermethylation of transcripts encoding factors involved in splicing, chromatin modification, and cell division could contribute to aberrant gene expression changes and contribute to the splicing, chromatin state and cell division defects previously observed upon EJC depletion, in addition to previously characterized pathways (Akhtar et al., 2019; Blazquez et al., 2018; Boehm et al., 2018; Mao et al., 2016; Roignant and Treisman, 2010; Wang et al., 2014c).

EJC haploinsufficiency within neural stem cells in developing mice has been shown to cause neurodevelopmental defects, which may model neurodevelopmental disorders in humans that are caused by EJC mutations (McMahon et al., 2016). p53, a central regulator of apoptosis and cell cycle progression, is upregulated in EJC haploinsufficient cells and mediates the observed neurodevelopmental defects (Mao et al., 2016). However, the molecular mechanism for the EJC-mediated regulation of p53 levels is unclear. We found that *TP53* mRNA, which encodes p53, is normally unmethylated but gains aberrant methylation upon *EIF4A3* KD (Figure 16E). Further, our differential expression analysis revealed that the upregulation of *TP53* mRNA upon *EIF4A3* KD observed in WT cells is abolished in *METTL3*$^{-/mut}$ cells (Figure 16D). p53 upregulation upon EJC KD is also disrupted in *METTL3*$^{-/mut}$ cells at the protein level (Figure 16F); note that the faster migrating band corresponds to an alternatively spliced, smaller p53 isoform that was previously reported to appear upon EJC depletion (Lu et al., 2017). These results indicate that eIF4A3 downregulates p53 expression through its effects on m^{6}A. Altogether, these findings reveal regulation of m^{6}A specificity as a previously unrecognized pathway by which the EJC impacts gene expression.

Physiological methylation of EJC-suppressed regions on transcript variants with long exons

To assess whether the EJC-mediated m^{6}A suppression is functionally relevant *in vivo*, we investigated whether unmethylated transcript regions that contain EJC-suppressed regions in HeLa cells were also suppressed in 25 different human tissues for which transcriptome-wide m^{6}A profiles are available (Liu et al., 2020b). We found that the vast majority (>95%) of these EJC-suppressed regions were also suppressed *in vivo* (Figure 17). Most regions were not methylated in any tissue, indicating that EJCs ubiquitously suppress many m^{6}A sites in a tissue-independent

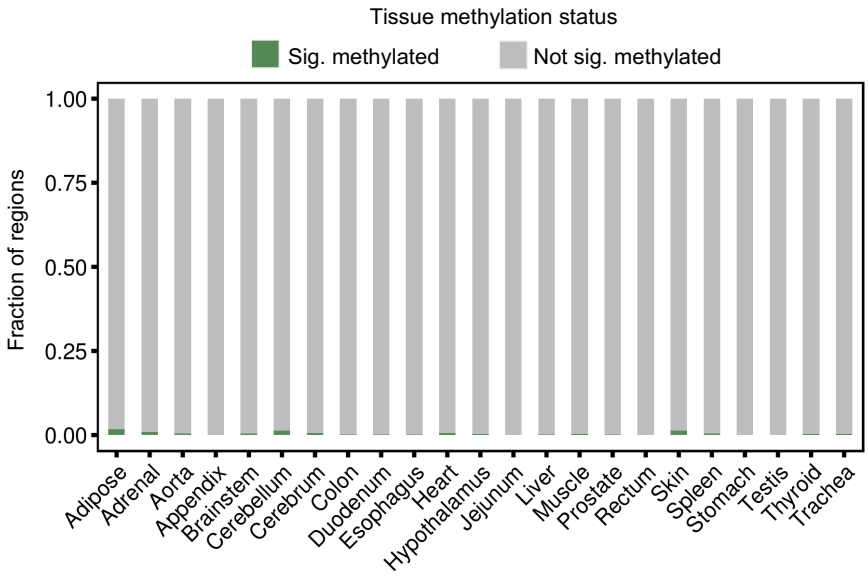

Figure 17. Most EJC-suppressed m^{6}A regions in HeLa cells are ubiquitously suppressed *in vivo* across human tissues.

Fractions of EJC-suppressed m^{6}A regions that are significantly methylated or not significantly methylated in different human tissues.

manner. This result is consistent with the fact that the global enrichment of m^{6}A in long internal exons and near stop codons is conserved across human tissue types (Liu et al., 2020b). These results suggest that EJC controls global m^{6}A epitranscriptome specificity *in vivo* across human tissues.

While the vast majority of the thousands of EJC-suppressed regions appear ubiquitously suppressed, we noted that hundreds of regions escape suppression in various tissues (Figure 18A). Notably, the methylation status of these regions varies across tissues, suggesting that these regions correspond to tissue-specific m^{6}A peaks (Figure 18B). We sought to understand the mechanisms by which these m^{6}A peaks escape EJC suppression in certain tissues. We hypothesized that some of these events may be explained by methylation of alternative transcript isoforms with altered exon structure and EJC binding locations. Indeed, we find that 100 of the tissue-specific m^{6}A peaks that escape EJC suppression span exon-intron boundaries and have significant m^{6}A-IP read coverage in regions annotated as intronic in the canonical transcript, as depicted for *SGSM2, FAM120A* and *MBTPS1* as examples *(*Figure 18C and Table 3). We inspected the aligned m^{6}A-IP reads and observed that these m^{6}A peaks contain reads that span the canonical exon-intron junction. These junction-spanning reads represent RNA isoforms that do not use that splice site, and thus presumably lack EJC binding at that location. This suggests that methylation of alternative transcript isoforms mediates the differences in m^{6}A methylation across tissues. We quantitatively assessed across tissues the correlation between the number of m^{6}A-IP reads that span the canonical exon-intron junction to the m^{6}A enrichment at the corresponding peak region in each tissue. Indeed, we found that the number of reads that span the junction in each tissue can explain a significant proportion of variation in m^{6}A enrichment across tissues (Figure 18D). Therefore, these regions, which are ordinarily suppressed by the EJC due to their close proximity to flanking

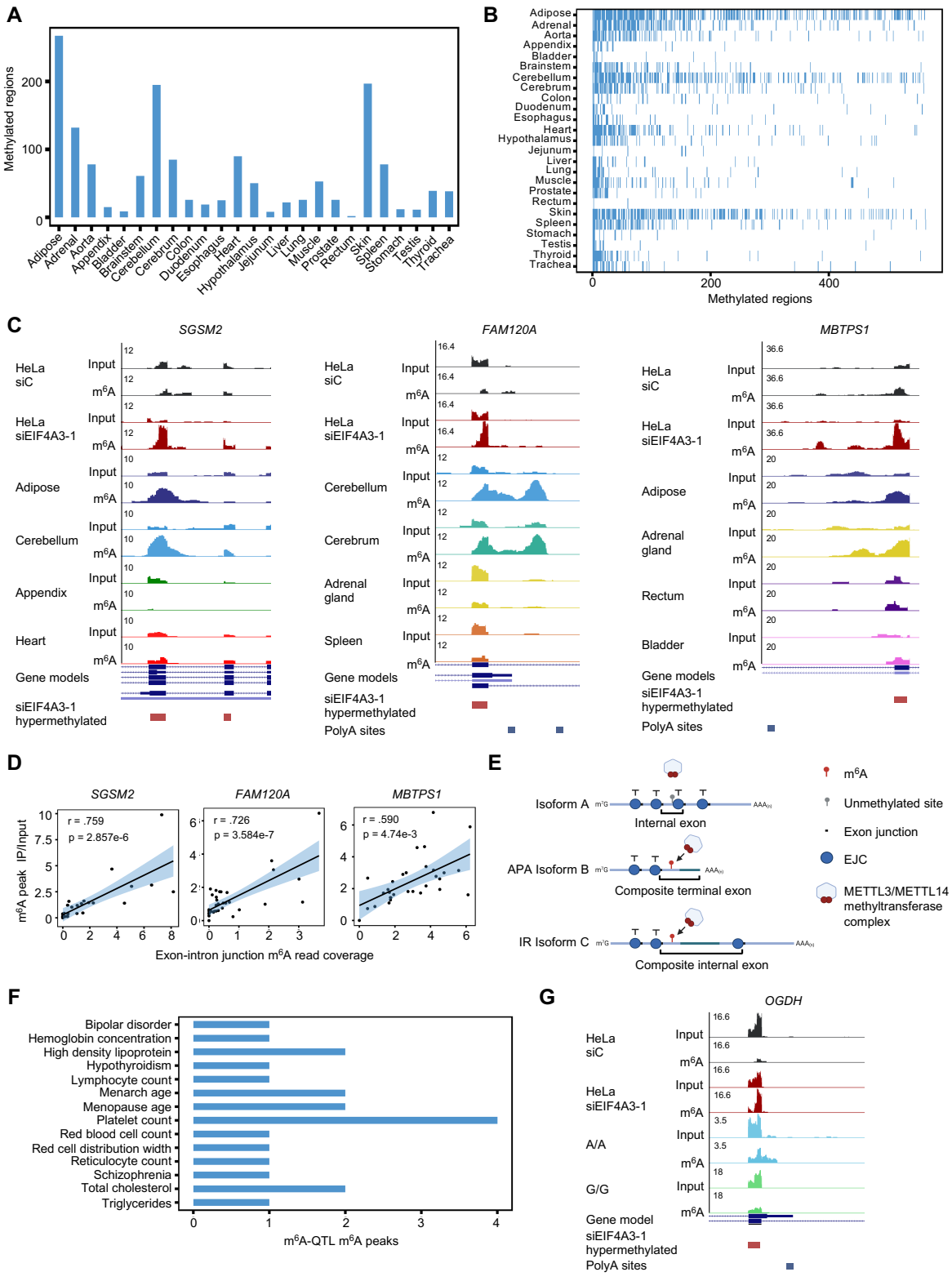

Figure 18. Suppressed m^{6}A sites physiologically escape exon junction complex silencing and mediate methylation variation across tissues and individuals.

Figure 18, continued.

A, eIF4A3-suppressed regions that are unmethylated in HeLa cells that are methylated in different human tissues. **B,** Methylation status of eIF4A3-suppressed regions across different human tissues. **C,** Input and m^{6}A-IP read coverage in HeLa siC, HeLa siEIF4A3-1 and representative tissues for *SGSM2, FAM120A* and *MBTPS1,* and gene models and annotated polyadenylation sites. **D,** Correlations between exon-intron junction m^{6}A read coverage and m^{6}A peak IP/input enrichment ratio, Pearson correlation. **E,** Schematic model for suppression of methylation within the internal exon of the canonical isoform (Isoform A), escape from suppression at the corresponding site in an isoform in which the site resides in a longer composite terminal exon due to intronic polyadenylation (APA Isoform B), and escape from suppression at the corresponding site in an isoform in which the site resides in a longer composite internal exon due to intron retention (IR Isoform C). **F,** Numbers of m^{6}A-QTL m^{6}A peaks that overlap eIF4A3-suppressed regions in Yoruba lymphoblastoid cell lines that are significantly associated with specified complex traits. **G,** Input and m^{6}A-IP read coverage in HeLa siC, HeLa siEIF4A3-1 and Yoruba lymphoblastoid cell lines NA18498 and NA18507. NA18498 is genotype A/A at the *OGDH* m^{6}A-QTL SNP (rs740040). NA18507 is genotype G/G. Gene models and annotated polyadenylation sites are shown below.

upstream and downstream EJCs, appear to physiologically escape m^{6}A suppression due to the lack of splicing and EJC deposition on the unspliced exon-intron boundary.

We next probed the identities of these alternative transcript isoforms. We analyzed transcriptome-wide intron retention in tissues with IRFinder and found that 53 of these exon-intron spanning m^{6}A peaks reside on alternatively spliced, retained introns, including the *SGSM2* m^{6}A peak (Fig. 18C and Table 4) (Middleton et al., 2017). We also noted that some of the peaks appear to occur on alternative isoforms that are generated through intronic alternative cleavage and polyadenylation (APA), illustrated by *FAM120A* and *MBTPS1* (Fig. 18C). Both these alternative pre-mRNA processing events alter transcript exon structure by lengthening exons. Intron retention at these sites converts an average-length internal exon into a long composite internal exon that contains two exons and their intervening intronic sequence. Similarly, APA at these sites converts an average-length internal exon into a longer composite terminal exon that contains both the internal exon and downstream intronic sequence. This expansion of exon length and removal of a flanking EJC appears to allow the suppressed m^{6}A sites within the internal exon to escape EJC suppression and acquire methylation (Figure 18E). This model is consistent with the results from our *CRY1* minigene experiments, in which we demonstrated that removal of either an upstream or downstream intron is sufficient for m^{6}A in an average-length internal exon to escape suppression (Figure 10A). A variety of studies have demonstrated that intron retention and intronic alternative polyadenylation play pivotal roles in normal and disease-related human biology (Lee et al., 2018; Naro et al., 2017; Wong et al., 2013). Our results suggest that a subset of EJC-suppressed regions can physiologically escape suppression through methylation of transcripts with longer exons and generate tissue-specific m^{6}A patterns that affect tissue-specific gene expression.

Furthermore, we similarly found that a subset of EJC-suppressed regions also escape silencing in Yoruba lymphoblastoid cell lines (LCL) that have been used to assess the impact of genetic variation on m^{6}A and other gene expression processes. A previous study identified m^{6}A peaks in LCLs that vary in methylation levels due to interindividual genetic variation (Zhang et al., 2020). We found that 48 of these variable peaks occur at EJC-suppressed regions in HeLa cells. Notably, variations in the m^{6}A levels of a subset of these peaks are significantly associated with complex traits, suggesting their functional importance (Figure 10F and Table 5). These traits include HDL, platelet and cholesterol levels, as well as disorders such as schizophrenia, hypothyroidism and bipolar disorder. For instance, genetic-mediated variation in methylation levels of an m^{6}A peak on *OGDH*, which encodes Oxoglutarate Dehydrogenase, is associated with variation in platelet counts. A SNP (rs740040) within an intron upstream of the *OGDH* m^{6}A peak is significantly associated with methylation level variation, with allele "A" conferring higher methylation levels; methylation variation in this transcript region appears to be driven by methylation of an APA transcript isoform, pointing to the functional relevance of this mode of m^{6}A regulation (Figure 10G). We note that many EJC-suppressed m^{6}A peaks in both tissues and LCLs can also escape suppression without apparent changes in transcript structure, suggesting the presence of additional pathways to enable methylation of these regions. Altogether, these results indicate that a subset of EJC-suppressed m^{6}A sites can escape suppression in certain physiological contexts, which contributes to methylation variation across tissues and individuals.

EJCs and peripheral EJC factor RNPS1 protect exon junction-proximal RNA regions from aberrant mRNA processing

We next investigated whether the EJC protects proximal regions from methylation due to a specific inhibitory interaction with the m^6A methyltransferase complex or due to steric hindrance. We immunoprecipitated METTL3-METTL14 and did not observe co-immunoprecipitation of EJC components, and immunoprecipitated the EJC and did not observe co-immunoprecipitation of METTL3-METTL14 (Figure 19). We further did not find evidence for any interactions between METTL3-METTL14 and EJC components in previously reported proteomic datasets of the respective complexes (Singh et al., 2012; Yue et al., 2018). The lack of interaction suggests that steric hindrance, rather than a specific inhibitory interaction, accounts for the ability of the EJC to protect proximal regions from methylation. Nuclear EJCs associated with the peripheral EJC factor RNPS1 multimerize and associate with wide variety of SR and SR-like proteins to package and compact mRNA into higher-order, megadalton-scale messenger ribonucloproteins (mRNPs) that ensheath proximal RNA well beyond the canonical EJC deposition sites (Mabin et al., 2018; Metkar et al., 2018; Singh et al., 2012). While monomeric EJCs assembled *in vitro* only protect a ~8-10 nucleotide footprint of RNA at the immediate EJC binding site, EJCs isolated from cellular extracts can protect tens to hundreds of nucleotides of proximal RNA from *in vitro* nuclease digestion due to this cellular mRNA packaging function (Le Hir et al., 2000; Singh et al., 2012). To assess whether the mRNA packaging function of the EJC mediates suppression of proximal methylation, we isolated EJCs/EJC-bound RNA from cellular extracts, digested away physically accessible RNA with *in vitro* nuclease treatment, then measured m^6A levels on the EJC-protected RNA footprints. Consistent with previous reports, we observed the presence of long EJC-protected RNA footprints tens to hundreds of nucleotides in length

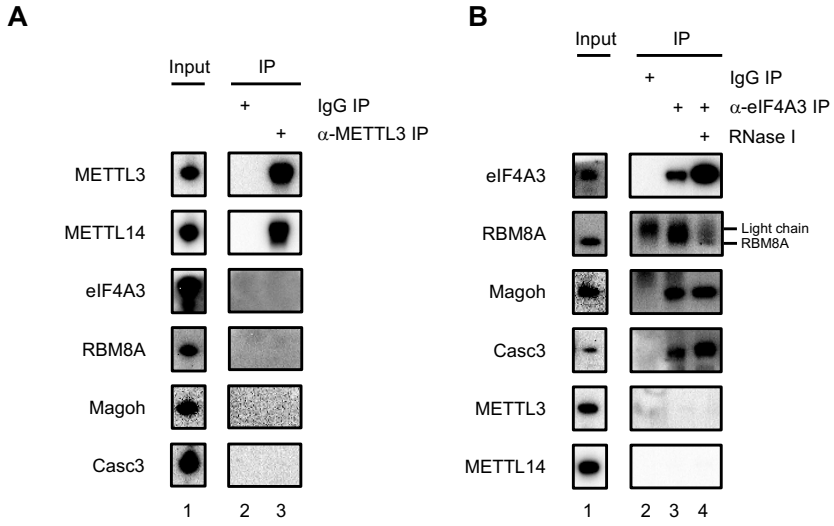

Figure 19. EJC and METTL3/METTL14 do not co-immunoprecipitate.
A,Western blots of EJC and METTL3/METTL14 proteins from total cell extracts (input) or upon immunoprecipitation with IgG or α–METTL3. **B,** Western blots of EJC and METTL3/METTL14 proteins from total cell extracts (input) or upon immunoprecipitation with IgG or α–eIF4A3, with or without RNase I digestion.

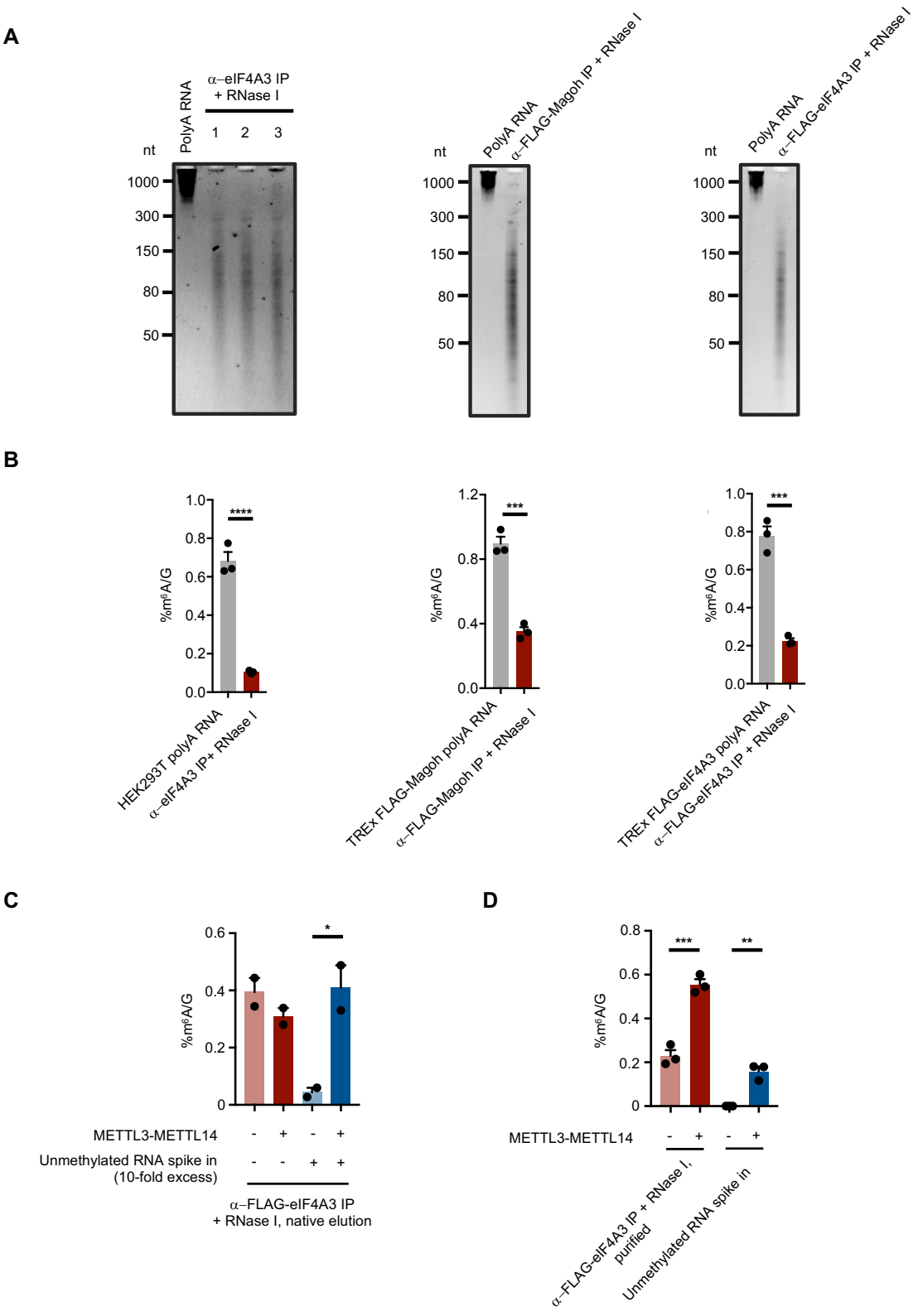

Figure 20. Exon junction complexes protect long RNA footprints from cellular m^{6}A deposition and in vitro m^{6}A deposition.

Figure 20, continued.
A, Length distributions of bulk polyadenylated RNA and RNase I-resistant EJC footprints. Length distribution of polyadenylated RNA (lane 1) and EJC footprints obtained from α–eIF4A3 immunoprecipitation and RNase I digestion in HEK293T cells (lane 2-4) (left). Length distribution of polyadenylated RNA (lane 1) and EJC footprints obtained from α–FLAG immunoprecipitation and RNase I digestion (lane 2) in TREx cells expressing FLAG-tagged Magoh (middle). Length distribution of polyadenylated RNA (lane 1) and EJC footprints obtained from α–FLAG immunoprecipitation and RNase I digestion (lane 2) in TREx cells expressing FLAG-tagged eIF4A3 (right). **B,** LC-MS/MS measurement of m^6A levels of bulk polyadenylated RNA and RNase I-resistant EJC footprints. m^6A levels of polyadenylated RNA and EJC footprints obtained from α–eIF4A3 immunoprecipitation and RNase I digestion in HEK293T cells (left). m^6A levels of polyadenylated RNA and EJC footprints obtained from α–FLAG immunoprecipitation and RNase I digestion in TREx cells expressing FLAG-tagged Magoh (middle). m^6A levels of polyadenylated RNA and EJC footprints obtained from α–FLAG immunoprecipitation and RNase I digestion in TREx cells expressing FLAG-tagged eIF4A3 (right). Three biological replicates, mean ± SEM , two-tailed T-test, ****P=.0002, ***P<.0005. **C,** LC-MS/MS measurement of m^6A levels of RNase I-resistant EJC footprints obtained by native elution of RNA-protein complexes and subsequent in vitro methylation with recombinant METTL3-METTL14. RNA-protein complexes were obtained by α–FLAG immunoprecipitation and RNase I digestion in TREx cells expressing FLAG-tagged eIF4A3, and then competitively eluted with FLAG peptide in methylation reaction buffer. For the right two bars, unmethylated RNA (in vitro transcribed MPm^6A library RNA) was spiked into the reactions in 10-fold excess to the amount of eluted EJC-bound RNA. Two biological replicates, mean ± SEM, two-tailed T-test, *P=.045. **D,** LC-MS/MS measurement of m^6A levels of Trizol-purified, deproteinized RNase I-resistant EJC footprints subjected to in vitro methylation with recombinant METTL3-METTL14 with the same reaction conditions used in c. Three biological replicates, mean ± SEM, two-tailed T-test, ***P=.0008, **P=.0016.

(Figure 20A). EJC-protected footprints are strongly depleted of m^6A, indicating that these inaccessible RNA regions are largely protected from m^6A deposition within cells (Figure 20B). Further, we found that EJCs protect these footprints from *in vitro* methylation by recombinant METTL3-METTL14 (Figure 20C). This is not simply due to general inhibition of methyltransferase activity, as free, unmethylated RNA spiked into the methylation reaction can be robustly methylated (Figure 20C). This is also not due to lack of methylatable sites on the EJC footprints, as deproteinized footprints are robustly methylated (Figure 20D). Altogether, these results indicate that EJCs suppress local m^6A deposition by packaging proximal RNA.

We next asked whether the peripheral EJC factor RNPS1, which associates with high molecular weight EJCs in these highly packaged mRNP structures, plays a role in protecting EJC-proximal regions from methylation (Mabin et al., 2018). We knocked down RNPS1 and, similarly to knockdown of core EJC factors, observed substantial hypermethylation within average-length internal exons and CDS regions (Figure 21A, 21B and 22A-D). We detected fewer hypermethylated regions overall compared to depletion of the core EJC factors; however, we observed high overlap between siRNPS1 hypermethylated regions and siEIF4A3/siRBM8A hypermethylated regions. Around 45% of significantly hypermethylated regions (628/1423) upon siRNPS1 KD correspond to regions that are hypermethylated upon depletion of core EJC factors, as depicted for *CDH24*, *NRAS,* and *CDCA7* as examples (Figure 21B and 22D). In contrast to the effects of RNPS1 depletion, depletion of UPF1, a central NMD factor that interacts with the EJC in the cytoplasm, did not result in m^6A changes similar to those of the core EJC (Figure 23).

If EJCs sterically hinder accessibility of proximal RNA, it follows that activities of other mRNA regulatory machineries may also be locally suppressed due to a general lack of accessibility. Notably, the ability of EJCs to protect proximal RNA regions from methylation

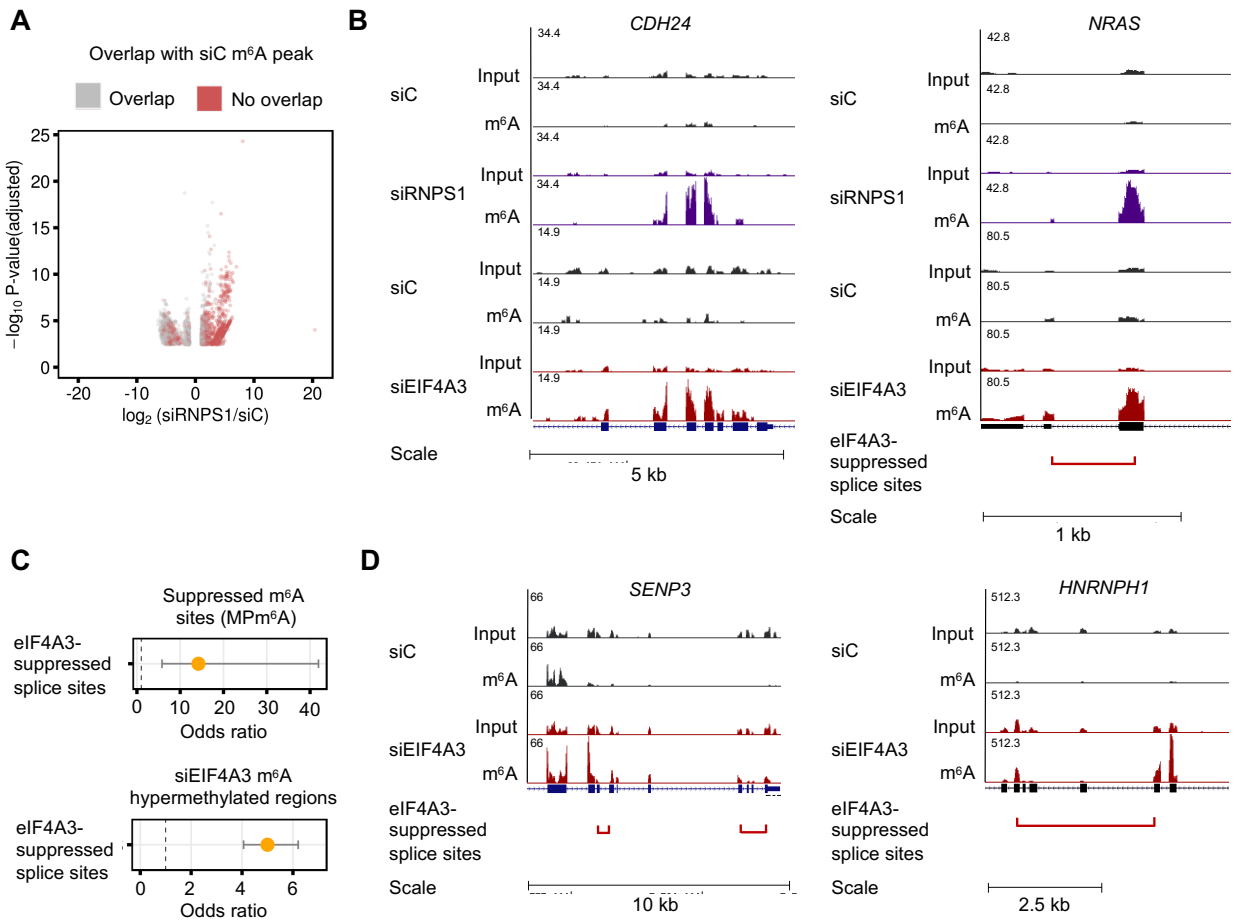

Figure 21. EJCs collaborate with peripheral factor RNPS1 to broadly protect proximal RNA regions from aberrant mRNA processing.

A, Differentially methylated regions upon *RNPS1* knockdown (KD) in HeLa cells (FDR<.1, $|\log_2 FC|>1$), three biological replicates. Gray dots indicate differentially methylated regions that overlap m^{6}A peaks in the non-targeting control KD cells, red dots indicate differentially methylated regions that do not overlap m^{6}A peaks in the control KD. **B,** Input and m^{6}A-IP read coverage at *CDH24* and *NRAS* upon *RNPS1 KD* and *EIF4A3* KD and corresponding non-targeting control KD in HeLa cells. Red bracket indicates an eIF4A3-suppressed splice variant, with ends of bracket indicating the suppressed splice junctions. **C,** Enrichment of suppressed splice sites activated upon *EIF4A3* KD at MPm^{6}A suppressed CDS m^{6}A sites relative to MPm^{6}A endogenous CDS m^{6}A sites (P=1.298e-12) (top) and at siEIF4A3-1 hypermethylated regions from HeLa relative to m^{6}A peaks in HeLa that are not hypermethylated (P=5.531e-70) (bottom). Fisher's exact test, dot and bar represent odds ratio and 95% confidence interval. **D,** Input and m^{6}A-IP read coverage at *SENP3* and *HNRNPH1* upon *EIF4A3* KD and non-targeting control KD in HeLa cells. Red bracket indicates an eIF4A3-suppressed splice variant, with ends of bracket indicating the suppressed splice junctions.

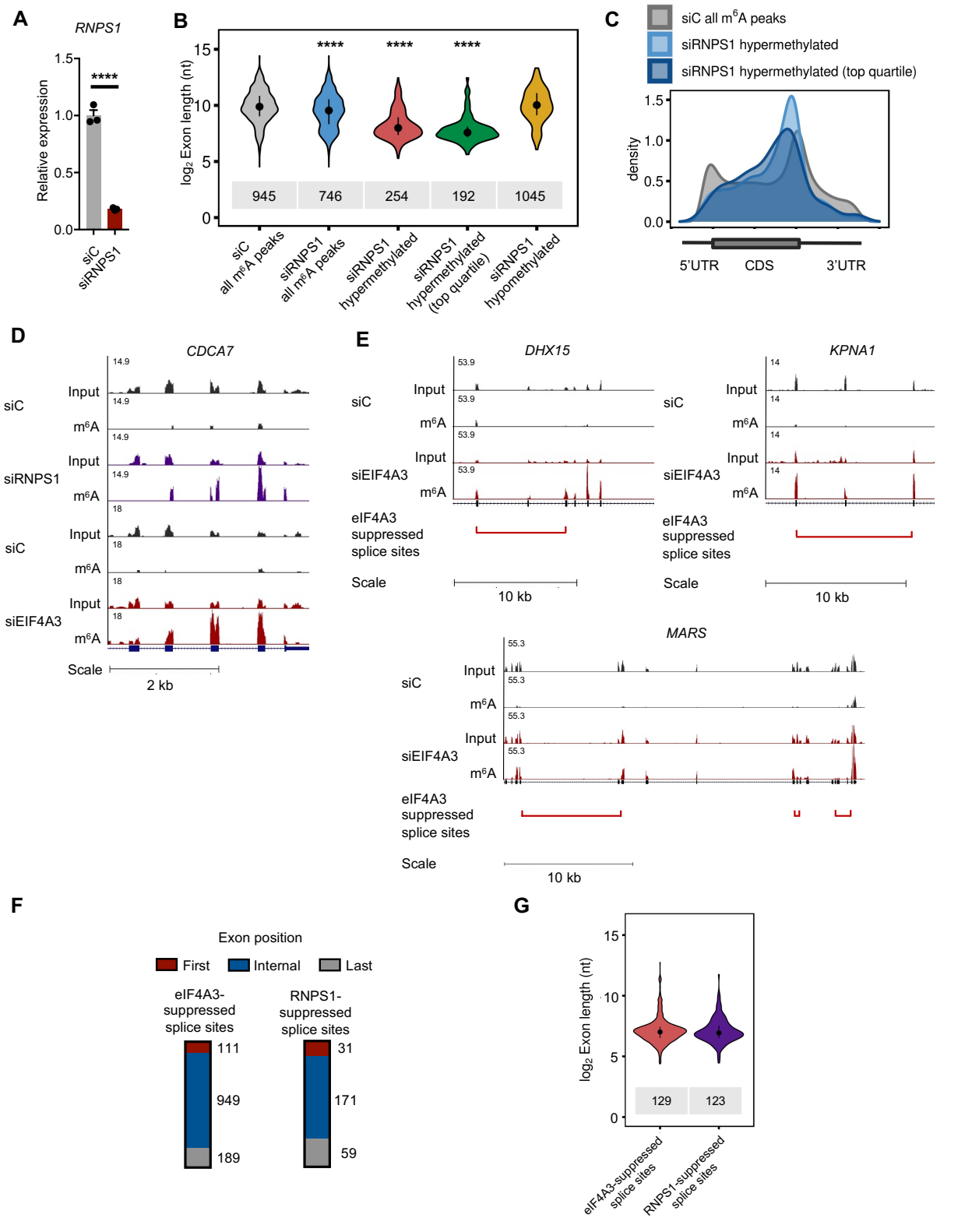

Figure 22. EJCs collaborate with peripheral factor RNPS1 to broadly protect proximal RNA regions from aberrant mRNA processing, continued.

Figure 22, continued.
A, *RNPS1* knockdown efficiency, normalized to siC, mean ± SEM, two-tailed T-test, ****P<.0001. **B,** Exon lengths of all m^6A peaks residing within internal exons in control KD and *RNPS1* KD, and m^6A hypermethylated and hypomethylated regions within internal exons upon *RNPS1* KD in HeLa cells. Dot and bar represent median and interquartile range, Wilcoxon rank sum test of indicated group vs. siC all m^6A peaks, ****P<2.2e-16. **C,** Metagenes of all m^6A peaks in control KD, significantly m^6A hypermethylated regions, and the top quartile of significantly m^6A hypermethylated regions upon *RNPS1* KD. **D,** Input and m^6A-IP read coverage at *CDH24* and *CDCA7* upon *RNPS1* KD, *EIF4A3* KD and non-targeting control KD in HeLa cells. **E,** Input and m^6A-IP read coverage at *DHX15*, *KPNA1* and *MARS* upon *EIF4A3* KD and non-targeting control KD in HeLa cells. Red bracket indicates an eIF4A3-suppressed splice variant, with ends of bracket indicating the suppressed splice junctions. **F,** Number of eIF4A3-suppressed and RNPS1-suppressed splice sites that unambiguously reside within first, internal or last exons. **G,** Exon lengths of eIF4A3-suppressed (median = 129 nt) and RNPS1-suppressed splice sites (median =123 nt) that reside in internal exons. Dot and bar represent median and interquartile range.

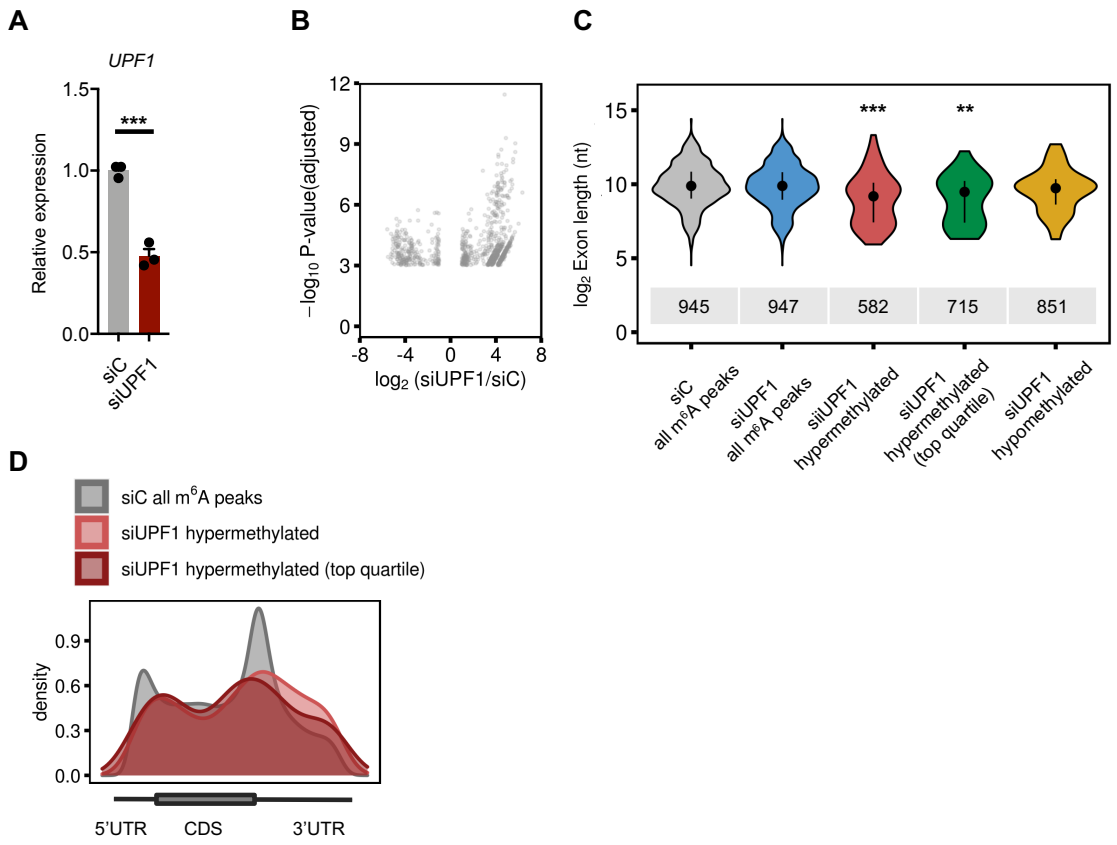

Figure 23. Nonsense-mediated decay factor UPF1 does not preferentially suppress m^{6}A methylation within average-length internal exons in mRNA coding sequences.

A, *UPF1* knockdown efficiency, normalized to siC, mean ± SEM , two-tailed T-test, ***P<.001. **B,** Differentially methylated regions upon *UPF1* KD in HeLa cells (FDR<.1, |log$_2$FC|>1), three biological replicates. **C,** Exon lengths of all m^{6}A peaks within internal exons in control and upon *UPF1* KD, and exon lengths of m^{6}A hypermethylated and hypomethylated regions within internal exons upon *UPF1* KD. Dot and bar represent median and interquartile range, Wilcoxon rank sum test of specified group vs. siC all m^{6}A peaks, ***P=1.815e-7, **P=.00654. **D,** Metagenes of all m^{6}A peaks in control KD, significantly m^{6}A hypermethylated regions, and the top quartile of significantly m^{6}A hypermethylated regions upon *UPF1* KD.

resembles the recently characterized EJC- and RNPS1-mediated suppression of proximal aberrant splice sites and recursive splicing (Blazquez et al., 2018; Boehm et al., 2018). Similar to what we observed for m^{6}A, the strength of this protective effect decreases proportionally with distance from the EJC/exon junction. We noted that the region containing EJC/RNPS1-suppressed m^{6}A methylation EJC-suppressed splice sites in *NRAS* coincides with EJC-suppressed splice sites (Figure 21B)(Boehm et al., 2018). Transcriptome-wide, EJC-suppressed m^{6}A sites significantly colocalize with EJC-suppressed m^{6}A sites, representing common regions in which both the m^{6}A methylation and splicing machineries are suppressed by the EJC (Figure 21C). Overall, we identify 592 eIF4A3-suppressed and 125 RNPS1-suppressed splice sites that reside within 100 nt of hypermethylated m^{6}A regions with siEIF4A3 KD, with representative examples shown at *SENP3*, *HNRNPH1*, *DHX15*, *KPNA1* and *MARS* (Figure 21D, 22E and Table 6). Consistent with the co-localization of suppressed methylation and splice sites, suppressed splice sites are predominantly found in average-length internal exons (Figure 22F and 22G). Altogether, these results suggest that RNPS1-associated EJCs suppress both local cellular m^{6}A methylation and splicing through packaging of proximal RNA and point to exon architecture as an important determinant of local RNA accessibility to regulatory machineries.

If long internal exons are broadly more accessible than average-length internal exons, one would predict that other mRNA regulators beyond the m^{6}A methyltransferase complex may also prefer to bind long internal exons. We performed a systematic analysis of whether other RBPs show a binding preference for long internal exons, using existing eCLIP data (Van Nostrand et al., 2020). Interestingly, in addition to RBM15, a subset of profiled RBPs also preferentially bind to long internal exons (Figure 24A). For instance, *SPEN* mRNA contains an 8,176 nt internal exon that is highly methylated and is also bound by RBPs that prefer binding on long internal exons

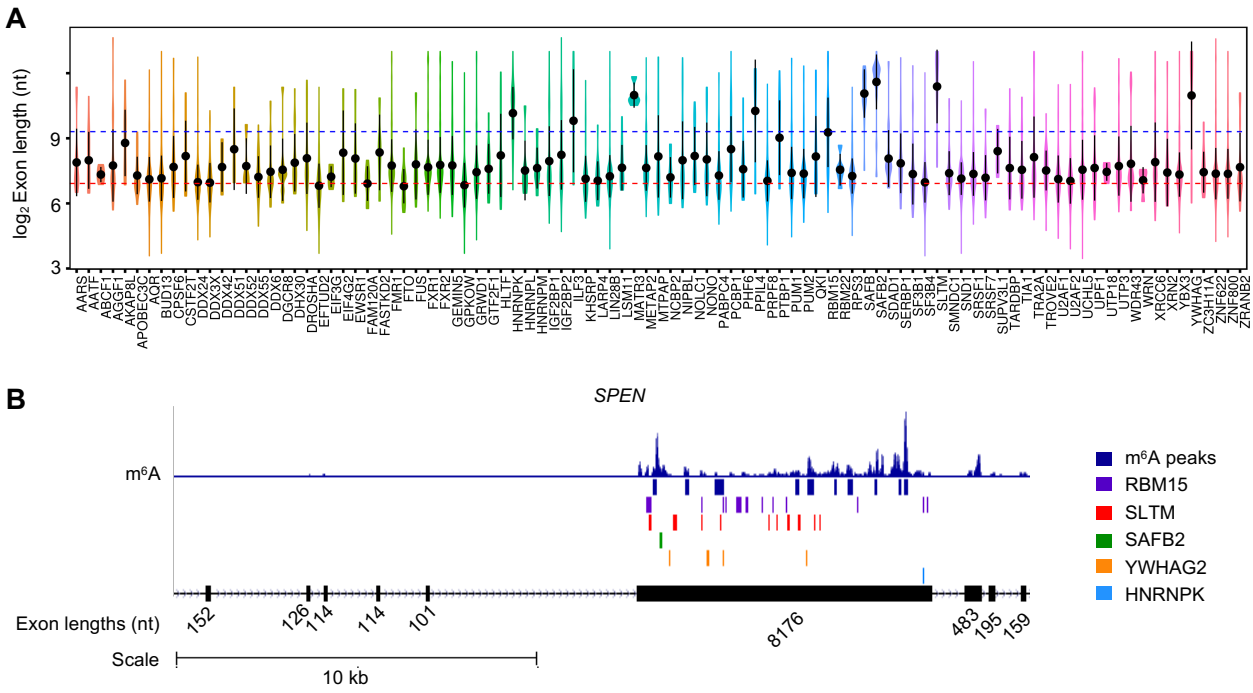

Figure 24. A subset of RBPs preferentially bind long internal exons.

A, Lengths of internal exons bound by 97 RBPs in K562 cells. Red line indicates the median exon size of all expressed internal exons in K562 cells (122 nt), blue line indicates the median exon size for internal exons bound by RBM15, a m^{6}A methyltransferase accessory subunit (626 nt). Dot and bar represent mean ± SD. **B,** m^{6}A methylation and RBP binding sites on SPEN mRNA.

(Figure 24B). To ensure that this length preference does not simply reflect the fact that RBP binding sites are more likely to fall in long internal exons by chance due to their greater lengths relative to short internal exons, we also calculated the ratio of RBP binding to longer (>200nt) vs. shorter (<200nt) internal exons and normalized by the number of nucleotides in each group. We still observed a strong preference for long internal exons for this subset of RBPs (Table 7). These observations suggest that EJCs may regulate mRNA accessibility to a broader range of mRNA regulators, in addition to the splicing and m^6A methylation machineries, through its mRNA packaging function (Figure 25).

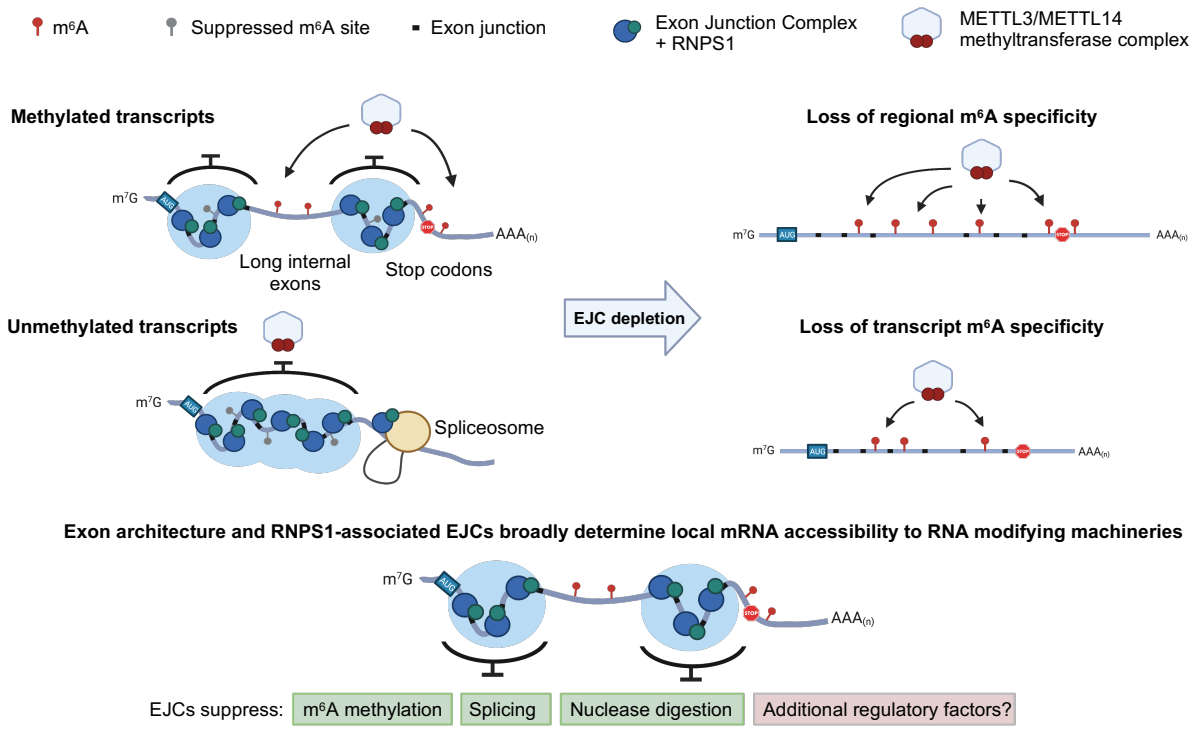

Figure 25. Exon junction complexes protect proximal RNA from methylation and other mRNA modifying processes by packaging mRNA.

Schematic model depicting RNPS1-associated exon junction complexes as widespread protectors of exon junction-proximal RNA from m^{6}A methylation, and other mRNA modifying activities, through local mRNA packaging.

Supplementary table legends

Table 2. Knockdown efficiency of *EIF4A3, RBM8A, RNPS1, UPF1* in m^{6}A-meRIP-seq experiments.
Supplementary Excel file listing the log_2FC in mRNA levels of targeted factors for all m^{6}A-meRIP-seq experiments.

Table 3. Tissue m^{6}A peaks that contain EJC-suppressed m^{6}A regions and span exon-intron boundaries.
Supplementary Excel file listing the genomic coordinates of tissue m^{6}A peaks that contain EJC-suppressed m^{6}A regions and span exon-intron boundaries

Table 4. Retained intron tissue m^{6}A peaks that contain EJC-suppressed m^{6}A regions.
Supplementary Excel file listing the genomic coordinates of retained intron tissue m^{6}A peaks that contain EJC-suppressed m^{6}A regions.

Table 5. m^{6}A-QTL-associated m^{6}A peaks that contain EJC-suppressed m^{6}A regions .
Supplementary Excel file listing the genomic coordinates of m^{6}A-QTL-associated m^{6}A peaks that contain EJC-suppressed m^{6}A regions, as well as the associated trait and m^{6}A-QTL SNP ID.

Table 6. eIF4A3-suppressed and RNPS1-suppressed splice sites in proximity to eIF4A3-suppressed m^{6}A regions.
Supplementary Excel file listing the genomic coordinates of eIF4A3-suppressed and RNPS1-suppressed splice sites that are within 100 nucleotides of eIF4A3-suppressed m^{6}A regions.

Table 7. Normalized ratio of RBP binding to >200nt and <200 nt internal exons.
Supplementary Excel file of the ratio of RBP binding to longer (>200nt) vs. shorter (<200nt) internal exons, normalized by the number of nucleotides in each group, for 97 RBPs with eCLIP data from K562 cells.

Table 8. Oligonucleotide sequences, antibodies, and other reagents.
Supplementary Excel file listing the oligonucleotide sequences, antibodies, and other reagents used in this study.

DISCUSSION

Overview

In this work, we developed and validated a massively parallel reporter assay that enables high throughput measurement of the m^{6}A methylation status of thousands of designed sequences in parallel, which we term Massively Parallel Assay for m^{6}A (MPm^{6}A). We then applied MPm^{6}A to systematically assess determinants of m^{6}A on an epitranscriptome-wide scale. We identified thousands of endogenously unmethylated DRACH sequences that are methylated when expressed outside of their native pre-mRNA context or when incubated *in vitro* with recombinant METTL3-METTL14 enzyme. These suppressed m^{6}A sites are concentrated in unmethylated transcriptome regions, revealing widespread suppression of m^{6}A as a major determinant of global m^{6}A specificity. We subsequently identified a mechanism that mediates a large portion of this suppression. We found that spliceosomes selectively suppress m^{6}A methylation in an exon length-dependent manner by depositing EJCs upstream exon junctions. EJCs protect long stretches of exon junction-proximal RNA from cellular methylation deposition by collaborating with the peripheral EJC factor RNPS1 to package and protect proximal RNA. We find that long EJC-bound RNA footprints are protected from methylation within cells and from in vitro methylation by recombinant METTL3/METTL14. This mechanism accounts for the suppression of many suppressed m^{6}A sites identified by MPm^{6}A and mediates multiple global characteristics of m^{6}A specificity. EJC protection of proximal RNA from methylation is essential for proper gene expression, as EJC depletion results in pervasive aberrant methylation of mRNAs and leads to m^{6}A-mediated gene expression dysregulation. While most EJC-suppressed m^{6}A sites appear ubiquitously suppressed, a subset of EJC-suppressed m^{6}A sites appear to escape silencing in

certain contexts due to methylation of transcript isoforms with longer exons that are produced by alternative splicing events. This contributes to methylation variation across tissues and inter-individual variation in m^{6}A associated with complex traits. Finally, suppressed m^{6}A sites co-localize with EJC/RNPS1-suppressed splice sites in average-length exons, suggesting that exon architecture broadly determines local RNA accessibility to mRNA regulatory machineries due to packaging of exon junction-proximal RNA by EJCs (Figure 26).

Suppression of m^{6}A methylation is a major determinant of m^{6}A epitranscriptome specificity

The extent to which m^{6}A levels across the transcriptome are controlled by intrinsic determinants vs extrinsic determinants has been the subject of debate. Using MPm^{6}A, we discovered thousands of potential m^{6}A sites in the transcriptome that are methylated when expressed outside of their native pre-mRNA context. This clearly indicates that local sequence features are not sufficient to encode endogenous methylation status and points to the importance of regulation of m^{6}A methylation of cellular mRNA by extrinsic factors. While an intrinsic preference of the METTL3/METTL14 methyltransferase complex for certain sequences is likely indeed one important determinant of m^{6}A specificity, it is not sufficient to explain patterns of m^{6}A deposition on cellular mRNA.

Previous studies investigating how m^{6}A is specifically deposited at specific sites on transcripts have mostly focused on mechanisms that selectively promote methylation at specific sites. Trans-acting factors such as transcription factors, RNA binding proteins, and histone modifications, have been proposed to recruit the methyltransferase complex to specific locations to preferentially methylate nearby RNAs (Barbieri et al., 2017; Bertero et al., 2018b; Fish et al., 2019; Huang et al., 2019; Slobodin et al., 2017b; Zhang et al., 2020). Outside of these recruitment

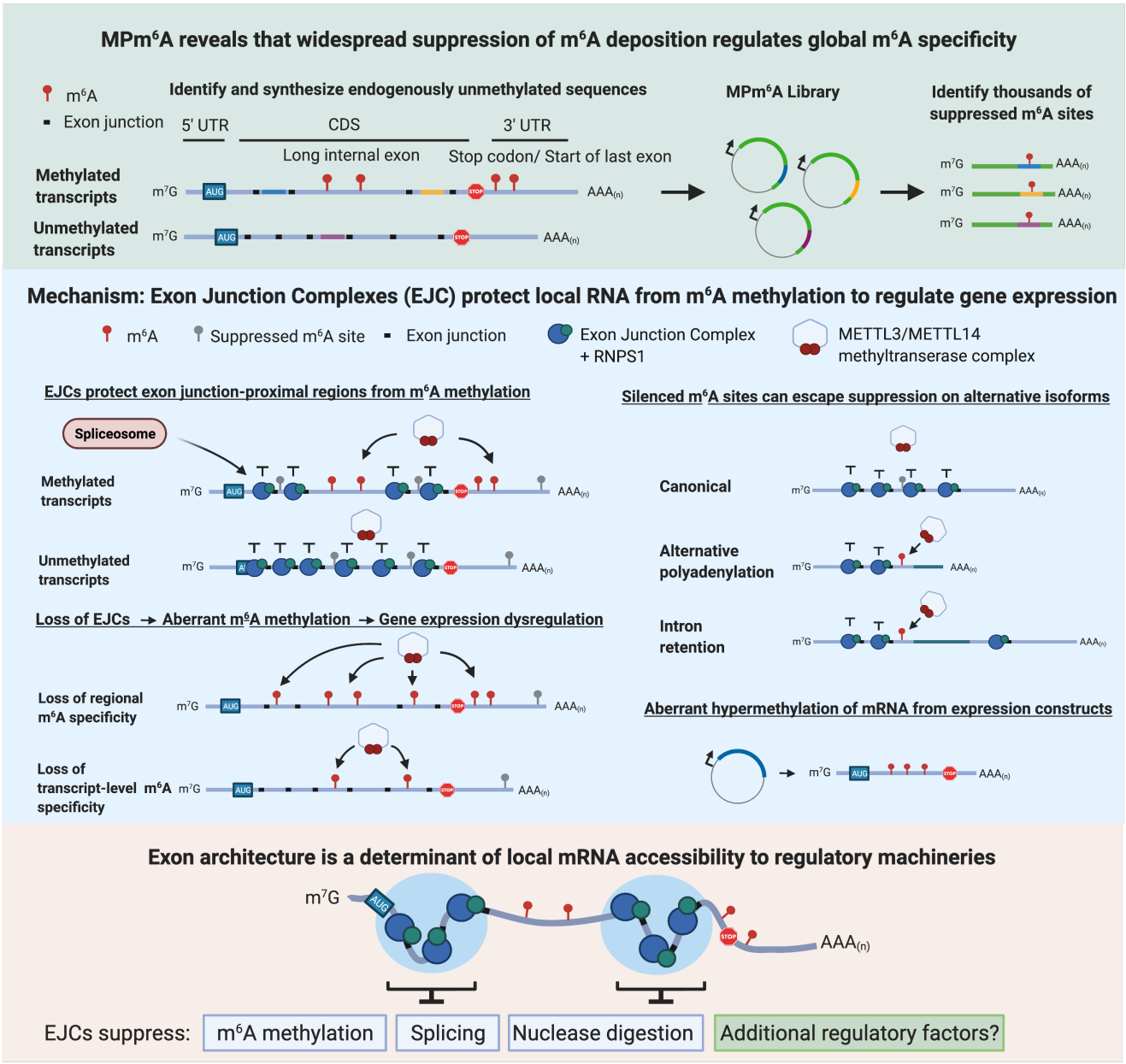

Figure 26. MPm^{6}A identifies exon junction complexes as m^{6}A suppressors.

Schematic model describing use of MPm^{6}A to discover that widespread suppression of m^{6}A controls global m^{6}A epitranscriptome specificity (top). Schematic model depicting the mechanism by which pre-mRNA splicing and subsequent exon junction complex deposition suppress local m^{6}A methylation (middle). Transcripts with altered exon architecture due to alternative pre-mRNA processing or expression from unspliced expression constructs can display altered m^{6}A patterns due to altered/absent EJC deposition by spliceosomes. Schematic model depicting exon

Figure 26, continued.
architecture as a broad determinant of local mRNA accessibility to RNA regulatory machineries via packaging of exon-junction proximal RNA by RNPS1-associated EJCs (bottom).

models, few alternative mechanisms that control selective m^{6}A deposition have been described. One previous study reported that transcription factor ZFP217 sequesters METTL3 to reduce global methylation levels in ESCs (Aguilo et al., 2015). However, it is not clear whether this mechanism impacts methylation selectivity. EJC protection of proximal RNA regions from methylation represents a fundamentally new mode of m^{6}A regulation that reveals the overlooked importance of suppressive mechanisms as major determinants of m^{6}A specificity.

The widespread existence of EJC-suppressed m^{6}A sites on unmethylated regions of transcripts, as well as transcripts that are ordinarily completely unmethylated, implies the existence of mechanisms that enable pervasive methylation of the transcriptome in the absence of EJC suppression. Known mechanisms that promote methylation, including interactions between METTL3/METTL14 and RNA polymerase II and interactions between METTL14 and H3K36me3, a histone mark that is found throughout active gene bodies across the genome, may facilitate widespread methylation of transcripts. This suggests that the m^{6}A landscape is shaped by a balance of activating mechanisms that enable widespread methylation of many transcript regions, and suppressive mechanisms, as characterized for EJCs, that confer methylation specificity by restricting methylation to certain regions. Our results do not exclude the possibility that significant portion of m^{6}A specificity can also be conferred by activating mechanisms, or that the stoichiometry of methylation sites may be modulated by activating mechanisms. However, our results do indicate that suppressive mechanisms are major regulators of m^{6}A deposition that mediate multiple characteristics of global m^{6}A specificity and are indispensable for safeguarding the transcriptome from otherwise widespread methylation.

Further, our systematic analysis of m^{6}A determinants with MPm^{6}A suggests the existence of more m^{6}A suppressors beyond the EJC. We identified thousands of suppressed m^{6}A sites in

3'UTRs, whose suppression appears to account for the decrease in m^{6}A enrichment past the stop codon. It has previously been reported that transcript isoforms that use proximal polyadenylation sites tend to be methylated and those that use distal polyadenylation sites tend to be unmethylated (Molinie et al., 2016). These sites are not suppressed by the EJC; m^{6}A suppressors that act preferentially on transcript isoforms that use distal polyadenylation sites may account for the silencing of these 3'UTR m^{6}A sites. Altogether, our results indicate that the EJC, and potentially other yet to be identified m^{6}A suppressors, are major regulators of m^{6}A specificity.

The Exon Junction Complex as an m^{6}A "suppressor"

Previously identified m^{6}A effector proteins fall broadly into three categories according to their activities: "writers", which catalyze m^{6}A methylation, "readers", which preferentially bind m^{6}A, and "erasers", which reverse m^{6}A methylation. Here we establish the EJC as a member of a new class of m^{6}A regulators: "suppressors", which broadly suppress the deposition of m^{6}A. The EJC appears to be a major regulator of m^{6}A deposition that controls multiple key aspects of global m^{6}A epitranscriptome specificity: the enrichment of m^{6}A in long internal exons, depletion of m^{6}A in CDSs and enrichment of m^{6}A in last exons near stop codons, and methylation selectivity for transcripts possessing long internal exons. This mechanism may also explain the high abundance of m^{6}A on chromatin-associated non-coding RNAs, such as enhancer RNAs and LINE-1 elements that are generally unspliced and thus not bound by the EJC (Liu et al., 2020a).

The EJC is known to play multifaceted roles in gene expression (Boehm and Gehring, 2016; Hir et al., 2016; Schlautmann and Gehring, 2020; Woodward et al., 2017). It promotes the nuclear export of spliced mRNAs by associating with the nuclear export adaptor ALYREF. It also plays a role in mRNA surveillance, marking transcripts with premature stop codons for NMD. It

has also become increasingly appreciated that NMD does not solely play a role in degrading "erroneous" transcripts, but also targets "normal" transcripts that contain stop codons upstream of last exons, such as those containing upstream open reading frames and 3'UTR introns for degradation. The EJC promotes translation by recruiting translation initiation factors, ribosome subunits and the ribosomal kinase S6K1 to mRNAs. The EJC controls splicing by blocking splicing at proximal cryptic splice sites. Further, the EJC has been reported to regulate transcription by stimulating promoter-proximal Pol II pausing by preventing the association of the transcription elongation factor P-TEFb complex with Pol II. This also impacts splicing, as promoter-proximal pausing prevents aberrant exon-skipping.

Our results indicate that the EJC, in addition to these characterized mechanisms, controls gene expression by suppressing m^6A methylation. For hundreds of genes, METTL3 depletion significantly diminished the upregulation or downregulation in mRNA levels normally observed upon *EIF4A3* depletion. This suggests that aberrant m^6A methylation resulting from EJC depletion exerts positive or negative effects on mRNA levels in a transcript-dependent manner. The heterogenous effects of m^6A methylation on mRNA levels observed here appear concordant with an emerging appreciation of the varied, context-dependent effects of m^6A on gene expression (Shi et al., 2019). As previously discussed, m^6A methylation can impact gene expression by a wide variety of mechanisms (He and He, 2021). It appears that no one mechanism is dominant enough across all methylation sites and across all biological contexts to allow one to predict the functional impact of m^6A methylation at any particular site *a priori*. For instance, a recent study identified m^6A-QTLs and observed low correlations in effect sizes between m^6A-QTLs and all other molecular QTLs analyzed, including expression QTLs, splicing QTLs, mRNA decay QTLs and protein QTLs (Zhang et al., 2020). One possible explanation for this result is that the functional

impact of m^6A on these processes is heterogenous due to context-dependent effects and diverse downstream mechanisms. Further work will be required to identify the contributions of various pathways through which aberrant methylation that is ordinarily suppressed by the EJC impacts gene expression. Additionally, given the widespread methylation changes observed, dissection of direct effects from indirect effects of methylation will also help elucidate the specific downstream pathways mediate the effects of aberrant m^6A methylation on gene expression. While further work will be needed to understand the downstream pathways, our results indicate that m^6A suppression is a previously unrecognized mechanism by which the EJC modulates gene expression.

EJCs exhibit several distinctive characteristics that endow them with a unique capacity to control gene expression. Unlike many other RBPs, EJCs bind RNA in a largely sequence-independent manner. Instead, their binding positions are specified by the splice site usage of a particular mRNA. Thus, EJCs to bind to all spliced mRNAs, regardless of sequence, which positions EJCs to impact the expression of most of the transcriptome. Further, most human genes are multi-exon, with relatively short internal exons and relatively long terminal exons. This stereotypical gene structure means that most mRNAs are bound by EJCs at closely spaced intervals within CDS regions. Further, EJCs are stably bound to mRNAs after their initial deposition, since the EIF4A3 subunit is a helicase that tightly clamps onto the RNA in an ATP-dependent manner. This may contribute to the ability of the EJC to package mRNA and protect it from accessibility, as the EJCs, once deposited onto mRNAs, do not dissociate from the RNA until the pioneer round of translation (Dostie and Dreyfuss, 2002; Lejeune et al., 2002). In conjunction with the widespread, stable binding of EJCs on most mRNAs, EJCs package and protect long stretches of proximal mRNA. These properties appear to collectively enable the EJC to robustly protect large

swaths of the transcriptome from methylation. Overall, we establish suppression of m^6A methylation as a novel function of the EJC that is important for gene expression.

Mechanistically, we propose that the EJC suppresses m^6A methylation of proximal mRNA via its mRNA packaging activity. EJCs have been reported to play key roles in the structural organization of pre-translational mRNPs by packaging and compacting RNA. EJCs initially form high molecular weight mRNP complexes in which EJCs multimerize and interact with a wide variety of other proteins, including SR and SR-like proteins such as the peripheral EJC component RNPS1 (Mabin et al., 2018). During later EJC stages, EJC composition shifts from this higher order mRNP complex to a lower molecular weight SR protein-devoid monomeric form that lacks RNPS1 and contains the EJC subunit CASC3. EJCs, in their high molecular weight form, have been proposed to package proximal RNA, based on their ability to protect long RNA footprints from *in vitro* nuclease digestion. However, the functional significance of EJC-mediated mRNP packaging has remained unclear.

We find that EJC-bound, long RNA footprints are protected from cellular methylation and in vitro methylation with recombinant METTL3-METT14. Further, RNPS1 depletion results in transcriptome-wide m^6A changes that resemble the changes observed upon depletion of the core EJC factors. This suggests that the ability of RNPS1-associated EJCs to package proximal mRNAs protects long stretches of proximal RNA from m^6A methylation. We do not detect a specific interaction between METTL3-METTL14 and the exon junction complex. The lack of interaction is concordant with the notion that steric hindrance, rather than a specific inhibitory interaction, accounts for the ability of the EJC to protect proximal regions from methylation.

Crosstalk between pre-mRNA splicing and m^{6}A methylation

Our results reveal widespread crosstalk between pre-mRNA splicing and m^{6}A methylation. m^{6}A has previously been reported to affect splicing (Liu et al., 2015; Louloupi et al., 2018; Xiao et al., 2016; Zhou et al., 2019a), and we now present evidence that, reciprocally, splicing is a widespread regulator of m^{6}A deposition. m^{6}A has been proposed to be deposited onto transcripts co-transcriptionally, though its general temporal relationship with splicing, another co-transcriptional process, is not entirely clear (Knuckles et al., 2017). Overall, the kinetics of m^{6}A deposition in relation to transcription and other RNA processing steps have not been characterized in detail. Based on m^{6}A-mapping of chromatin-associated and nascent RNA, METTL3/METTL14 appears to install methylation onto RNA soon after, or during the process of transcription (Ke et al., 2017). Methylation is often referred to as a co-transcriptional event, due to detected interactions of METTL3/METTL14 with RNA polymerase and chromatin (Knuckles et al., 2017; Slobodin et al., 2017a). However, the extent to which METTL3/METTL14 deposits m^{6}A onto RNA that is still tethered to the polymerase i.e, co-transcriptionally vs. at other early time points following release of the transcript does not appear to have yet been directly examined.

Reported roles for m^{6}A in splicing regulation indicate that some amount of m^{6}A is deposited onto pre-mRNA prior to splicing at certain sites (Louloupi et al., 2018; Xiao et al., 2016; Zhou et al., 2019b). Our results imply that a significant amount of m^{6}A is also deposited onto mRNA following splicing and that m^{6}A deposited onto pre-mRNA prior to splicing would not be expected to exhibit the hallmarks of m^{6}A specificity conferred by EJC suppression. Consistent with this notion, a study that mapped m^{6}A on nascent mRNA, which is enriched for unspliced pre-mRNA, reported increased m^{6}A enrichment within 100 nt of splice sites in nascent RNA when compared to polyadenylated mature mRNA (Louloupi et al., 2018). We analyzed this dataset and

found that m^{6}A peaks in nascent RNA are found in much shorter internal exons (median = 205 nt) than m^{6}A peaks in polyadenylated mRNA (938 nt), which appears to explain the increased enrichment near splice sites (Figure 27A). Moreover, nascent mRNA does not show the characteristic enrichment of m^{6}A near the stop codon observed in polyadenylated mRNA and exhibits a broader enrichment over coding sequences (Figure 27B). Thus, consistent with our finding that the EJC globally regulates m^{6}A specificity, the distribution m^{6}A in nascent mRNA differs markedly from normal polyadenylated mRNA and appears to parallel the polyadenylated distribution observed upon EJC depletion. This data further suggests that m^{6}A peaks that do reside in exon-junction proximal regions in mature mRNA may primarily be deposited on pre-mRNAs prior to splicing.

We note that one study mapped m^{6}A on chromatin-associated RNA and reported that it was similar overall to the distribution in nuclear and cytoplasmic RNA fractions; this result appears to be consistent with our results, as the chromatin-associated RNA (caRNA) analyzed in this particular study consists predominantly of spliced chromatin-associated RNA, with the majority of exons >90% spliced (Ke et al., 2017). We hypothesize that alterations in the relative kinetics of m^{6}A methylation and splicing could also mediate differential methylation of EJC-suppressed m^{6}A sites without generation of alternative transcript isoforms. Future work should further investigate the temporal order of splicing and m^{6}A methylation processes across the transcriptome.

Exon architecture as a functional element in gene expression regulation

Our results suggest that exon architecture is a functionally relevant element for post-transcriptional gene expression regulation. Human genes exhibit a stereotypical exon architecture of short internal exons of relatively uniform length (140-150 nt on average), flanked by longer

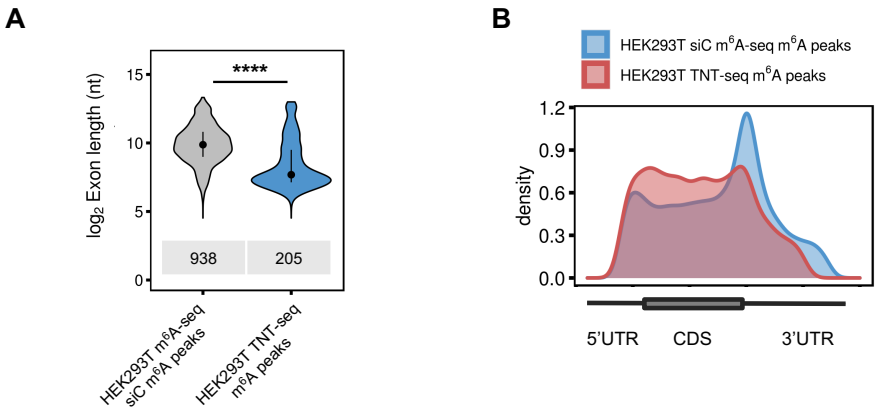

Figure 27. m^{6}A distribution in nascent RNA exhibits decreased enrichment in long internal exons and near stop codons.

A, Exon lengths of m^{6}A peaks that reside in internal exons in m^{6}A-meRIP-seq of bulk polyadenylated RNA and Transient N-6-methyladenosine transcriptome sequencing (TNT-seq) of nascent RNA in HEK293T cells. Dot and bar represent median and interquartile range, Wilcoxon rank sum test, **** = P<2.2e-16. **B,** Metagenes of m^{6}A-seq and TNT-seq peaks in HEK293T cells.

terminal exons. Whether this stereotypical exon architecture has functional importance for gene expression has not been extensively studied. One study reported that first exon length is inversely correlated with promoter-proximal H3K4me3 and H3K9ac levels and transcriptional activity (Bieberstein et al., 2012). The authors proposed that short first exons promote transcription because the decreased distance between the first 5' splice site and the promoter allows the splicing machinery to exert positive feedback on transcription. Overall, there is a paucity of data on the relevance of exon architecture and exon length on gene expression.

Here, we find that exon architecture is functionally relevant for gene expression as a principal determinant of mRNA m^6A methylation. Since EJCs are positioned near mRNA exon boundaries, exon architecture determines the spacing between EJCs. This spacing is relevant, as closely spaced EJCs can prevent methylation of long stretches of mRNA, whereas long gaps between EJCs permit methylation of the intervening sequence. For transcripts that exhibit a stereotypical exon architecture, this means that internal sequences that correspond to internal exons are largely protected from methylation by EJCs, while sequences in longer terminal exons can be methylated. Deviations from this stereotypical exon architecture result in deviations in methylation patterns; for instance, the minority of transcripts with unusually long internal exons are preferentially methylated over transcripts without long internal exons.

Notably, exon architecture is not fixed, but is regulated by alternative pre-mRNA processing. Alternative pre-mRNA processing events such as intron retention and intronic alternative cleavage and polyadenylation produce transcripts with altered exon architectures with longer exons (Lee et al., 2018; Monteuuis et al., 2019; Singh et al., 2018). We find that EJC-suppressed m^6A sites in average-length internal exons can be methylated on alternatively spliced transcript isoforms with longer exons, presumably due to the expansion of exon length resulting

in removal of a flanking EJC. Methylation of these transcript isoforms appears to contribute to variation in methylation across tissues and inter-individual variation in methylation associated with complex traits. The functional implications of differential methylation of transcript isoforms on gene expression should be further explored. We note that is currently unclear whether methylation is deposited onto these transcripts prior to or following the pre-mRNA processing steps that produce these transcripts. Some studies report that m^6A promotes alternative polyadenylation, though others report that it does not (Kasowitz et al., 2018; Ke et al., 2015; Molinie et al., 2016; Yue et al., 2018). Either case is consistent with the notion that lack of EJC binding enables suppressed m^6A sites to escape suppression and mediate differential methylation between transcript isoforms. The effects of exon length on mRNA modification could represent a previously unrecognized mode of gene expression control that is relevant in diverse physiological and pathophysiological processes.

Another setting in which exon architecture is commonly altered is when mRNA is expressed from artificial expression constructs. cDNA expression constructs lacking endogenous splice sites are widely used tools in many fields of biology and medicine. Importantly, we find that m^6A levels on mRNAs expressed from unspliced plasmids can significantly differ from endogenously spliced mRNAs due to activation of EJC-suppressed m^6A sites. This may impact the interpretation of experiments using these constructs. Given the pervasive presence of suppressed m^6A sites across the transcriptome, going forward, investigators studying m^6A methylation, and RNA regulation more generally, using unspliced plasmid-based constructs should check whether reporter mRNA recapitulates endogenous m^6A patterns to prevent artifactual findings. Aberrant m^6A deposition on these transcripts may also have relevance for other experimental or therapeutic contexts that in which cDNA expression constructs are used, such as

in gene therapy. For instance, we find that mRNA for *SMN1*, used in gene therapies for Spinal Muscular Atrophy, and *IDS*, used in gene therapies for Mucopolysaccharidosis II, is hypermethylated when expressed from cDNA constructs (AveXis, Inc., 2020; Scarpa, 1993). Altogether, our results point to the functional importance of exon architecture as a determinant of m^6A methylation and gene expression.

EJC-mediated mRNA packaging as a mechanism to broadly regulate mRNP accessibility

A particularly intriguing implication of our findings is that EJC-mediated mRNA packaging may serve as a broader system to control "mRNP accessibility". After demonstrating that EJCs sterically hinder methylation of proximal RNA, we investigated whether the activities of other mRNA regulatory machineries beyond the m^6A methyltransferase complex may also be locally suppressed by EJCs. We noted significant parallels between the ability of the EJC protect proximal RNA from methylation and the recently characterized ability of the EJC to protect proximal RNA from aberrant splicing (Blazquez et al., 2018; Boehm et al., 2018). Similar to what we observed for m^6A, the suppressive effect on splicing extends beyond the canonical EJC binding sites to nearby regions, and the strength of this protective effect decreases proportionally with distance from the EJC/exon junction. Both processes also involve the activity of the peripheral EJC component RNPS1. The magnitude of the protective effect is strong for both methylation and splicing, with many suppressed splice sites and m^6A sites exhibiting no detectable usage under normal conditions. Based on these observations, we propose that mRNP packaging by EJCs broadly limits the accessibility of exon junction-proximal RNA to RNA modifying machineries, thus protecting nearby RNA from methylation, splicing, and nuclease digestion. This would tie the

exon architecture of transcripts to the accessibility of specific mRNP regions to these regulatory machineries.

We note that despite these prominent similarities, there do appear to be several differences between the suppression of splicing and the suppression of methylation by EJCs. The range of EJC protection against methylation appears to be greater than that of EJC protection against splicing. Experiments in which EJC components are artificially tethered to minigene constructs indicate that the range of suppression for splicing is on the order of tens of nucleotides away from EJCs (Blazquez et al., 2018; Boehm et al., 2018). In contrast, our minigene experiments and EJC depletion experiments indicate that the range of m^6A suppression can extend from tens to hundreds of nucleotides away from the nearest EJCs. Further, in our siRNA knockdown experiments, we detect an order of magnitude more EJC-suppressed m^6A regions than previously reported EJC-suppressed splice sites. These results indicate that EJC suppression of m^6A may be relatively stronger than EJC suppression of splicing. However, we cannot exclude technical differences as a factor that contributes to the apparent differences in the properties of EJC-mediated m^6A and splicing suppression. Regardless, the similarities between these mechanisms, as well as the data presented here connecting EJC-mediated mRNA packaging to its m^6A suppression activity, suggest that transcript exon architecture may be an important determinant of local mRNP accessibility due to the mRNA packaging by EJCs.

Chromatin accessibility, defined as the degree to which nuclear macromolecules can physically contact chromatinized DNA, is well established as an important determinant of eukaryotic gene expression. Chromatin accessibility is a key determinant of transcription factor (TFs) binding specificities, which is essential for proper transcriptional regulation. Nucleosomes, which consist of an octamer of histone proteins encircled by ~ 147 bp of DNA, are the core

structural elements of chromatin and principal determinants of chromatin accessibility (Kornberg and Lorch, 1999). Nucleosomes physically occlude access of packaged DNA, broadly suppressing accessibility of bound DNA to DNA regulatory factors such as TFs. Chromatin that is loosely packaged, or "open" chromatin, is permissive for transcription factor binding, while chromatin that is tightly packaged, or "closed" chromatin, is not permissive for transcription factor binding. While there are certainly exceptions to this model, as evidenced by TFs that preferentially bind to heterochromatic DNA and "pioneer" transcription factors can bind to closed chromatin and participate in its conversion to open chromatin, most TFs whose genome-wide binding sites have been surveyed bind to accessible chromatin almost exclusively (Thurman et al., 2012).

In contrast to chromatin accessibility, the concept of mRNP accessibility as a determinant of gene expression is not as well established. mRNA is heavily associated with protein; mRNPs have been estimated to exhibit an RNA:protein weight ratio of around 1:3, calculated based on the equilibrium buoyant densities of mRNPs in a CsCl gradient (Perry and Kelley, 1966; Spirin, 1979). Pre-mRNA and mRNA are known to be bound by a complement of associated RBPs that varies between genes and changes throughout the lifetime of an mRNA (Singh et al., 2015). It is currently estimated that over 1000 proteins in human cells can bind to mRNA (Baltz et al., 2012; Bao et al., 2018; Queiroz et al., 2019; Trendel et al., 2019; Urdaneta et al., 2019). Some of these RBPs bind transiently, while others bind more stably and form large protein-RNA complexes that are more likely to impact general mRNP accessibility. Early reports describing RNA packaging focused on the packaging of nascent heterogeneous nuclear RNA (hnRNA), more commonly referred to as pre-mRNA today, by heterogenous nuclear ribonucleoproteins (hnRNP) into hnRNP particles. Interestingly, these reports revealed striking parallels between the packaging of hnRNA into hnRNP particles and the packaging of genomic DNA into nucleosomes. Visualization of newt

oocyte and *Drosophila* embryo hnRNA by electron microscopy revealed a "beads-on-a-string" repeating subunit structure of 200-300 angstrom spherical protein-associated particles connected by ribonuclease-sensitive regions(Malcolm and Sommerville, 1974; McKnight and Miller, 1976). These oligomeric arrays of ribonucleoprotein particles are converted to 20 nm monomers in the presence of nuclease activity. These monomers sediment near 40S in density gradients and are termed hnRNP particles. Various hnRNP proteins (hnRNPA1, hnRNPA2, hnRNPB1, hnRNPB2, hnRNPC1, hnRNPC2) were identified as the packaging proteins within these hnRNP particles in human cells (Beyer et al., 1977). These proteins assemble in a defined molar ratio and package pre-mRNA into repeating arrays (Barnett et al., 1990; Conway et al., 1988). hnRNPC1/C2 (hnRNPC) act as protein rulers that define the length of RNA to be packaged. hnRNPC forms tetrameric complexes that binds approximately 150-230 nt of RNA with high cooperativity (McAfee et al., 1996). Due to the notable parallels between the packaging of DNA by histones and the packaging of pre-mRNA by hnRNPs, these hnRNP particles have been previously been termed "ribonucleosomes", or "RNA nucleosomes" (Conway et al., 1988; König et al., 2010; Weighardt et al., 1996).

Relatively few studies appear to have examined how the pre-mRNA packaging function of hnRNPs specifically relate to their function of individual hnRNPs, as opposed to packaging-independent activities. One study examining the role of hnRNP packaging in gene expression mapped hnRNPC RNA binding sites and found that the RNA recognition motif domains of hnRNP C bind uridine-rich tracts (König et al., 2010). Consistent with the role of hnRNPC in packaging RNA in hnRNP particles, a subset of these binding sites are spaced at regular intervals of ~165 nt and 300 nt. This study proposed that hnRNP packaging plays a dual role in modulating alternative splicing. A subset of alternative exons appear to be silenced by incorporation of the exonic RNA

sequence into the hnRNP particles, whereas the usage of a subset of alternative exons are enhanced by incorporation of RNA in the preceding intron into hnRNP particles. Overall, while there is some evidence for functional roles of hnRNP particles in gene expression, more detailed investigation is needed to fully elucidate their significance.

EJCs have more recently been reported to exhibit the ability to package mRNA, representing a system by which mature mRNAs are packaged and compacted into mRNPs (Mabin et al., 2018; Metkar et al., 2018; Singh et al., 2012). There are intriguing similarities between EJC-mediated mRNP packaging to the systems of packaging that have been characterized on pre-mRNA and DNA, but also notable differences. EJCs share a notable similarity with histone octamers regarding their ability to extensively bind their respective nucleic acids at closely-spaced intervals. The average distance between nucleosome midpoints in humans is ~185 bp, and the average exon length (and thus the average spacing between EJC binding sites) is around 140 nt (Schwartz et al., 2009; Tilgner et al., 2009). While hnRNPs are reported to bind RNA at ~165 nt intervals, which resembles nucleosome and EJC spacing, it is unclear how regularly and extensively hnRNPC bind to each pre-mRNA given the stronger preference for specific sequence motifs than histones and EJCs. Like histones and hnRNPs, EJCs package and compact long stretches of their respective nucleic acid. However, the composition and structure of EJC-packaged mRNPs is not as well defined as nucleosomes and hnRNP particles. It is unclear whether the subcomponents of the high molecular weight EJC-packaged mRNP structures associate in a defined stoichiometry. Further, while EJCs bind mRNA at close intervals, it does not appear that they form simple monomeric particles resembling beads on a string (Metkar et al., 2018). Like hnRNP packaging of pre-mRNA, the functional roles of EJC packaging of mRNA have not been explored in great detail.

Our findings here, in conjunction with data from other recent reports, suggests that mRNP packaging plays an essential role in controlling accessibility of cellular RNA to the m^6A methylation and splicing machineries (Blazquez et al., 2018; Boehm et al., 2018). Our findings suggest that mRNP packaging could represent an important determinant of gene expression by broadly controlling the local accessibility of mRNA to regulatory machineries. We find that a subset of RBPs, like the m^6A methyltransferase complex, prefer to bind long internal exons over shorter internal exons. These RBPs are candidates for future investigation as factors whose specificity may also be controlled by EJC-mediated mRNP packaging. Future work should explore the extent to which mRNP accessibility represents an important general determinant of gene expression.

We note that the breadth of processes that are suppressed by the EJC may be limited by the temporal window in which mRNA resides in this tightly packaged state as well as by variations in EJC deposition. Since EJCs are thought to be displaced from mRNAs upon the pioneer round of translation, regulatory machineries that interact with mRNAs following translation are unlikely to be impacted by EJC-mediated mRNA packaging. Further, not all pre-translational events may be equally impacted by EJC-mediated packaging. As mentioned above, EJC mRNP composition is thought to shift from an initially high molecular weight, highly packaged form, to a lower molecular weight, loosely packaged form as the mRNA proceeds from the site of transcription in the nucleus to translation by ribosomes in the cytoplasm (Mabin et al., 2018). Pre-mRNA splicing and m^6A methylation are nuclear processes that generally occur quickly after transcription takes place, likely coinciding with the time when EJCs initially form highly packaged mRNP structures. Given that EJC composition may shift over time, events in mRNA metabolism that occur at later stages in the mRNA lifetime may evade EJC-suppression. We also note that transcriptome-wide

mapping of EJC binding sites indicates that while the majority (>80%) of exons exhibit EJC occupancy at the canonical EJC position, a subset do not exhibit EJC occupancy (Saulière et al., 2012; Singh et al., 2012). This may be in part due to RNA secondary structures at the canonical binding site that prevent EJC deposition, as well as other unknown factors. This also may be due to technical issues with methods used to map EJC binding sites that result in a certain frequency of false negatives, as genome-wide approaches often exhibit protocol-specific biases and imperfect sensitivity. For instance, it has been reported that EJCs do not photo-crosslink efficiently to RNA (Patton et al., 2020). Further work will be required to examine the breadth of molecular processes that are suppressed by EJC packaging of mRNPs.

Conclusion

Prior to this work, the mechanisms that globally control the selective installation of m^6A methylation on cellular mRNAs were not well understood. To assess the importance of intrinsic vs. extrinsic determinants of global m^6A specificity, we developed a novel massively parallel reporter assay approach that we term MPm^6A. MPm^6A enabled us to assess the methylation status of thousands of designed mRNA sequences in a high throughput manner. Using MPm^6A, we discovered that extrinsic factors that suppress thousands of potential methylation sites are essential for global control of m^6A specificity. Further investigation using a combination of bioinformatic analysis, minigene reporter constructs and siRNA knockdown revealed that many of these sites are suppressed by EJCs. EJCs mediate multiple global characteristics of m^6A specificity by packaging exon junction-proximal RNA following pre-mRNA splicing and protecting it from m^6A methylation. Since EJC positioning on mRNAs is determined by the exon structure that is created during pre-mRNA splicing, our findings indicate that exon architecture is a principal determinant

of m^{6}A methylation. For instance, some transcript variants with longer exons, generated through non-canonical splicing events such as IR and intronic APA, appear to be methylated in mRNA regions in which methylation is ordinarily suppressed by EJCs. Depletion of EJCs results in aberrant methylation of thousands of transcripts, leading to m^{6}A-mediated gene expression dysregulation. Finally, we note striking parallels between the ability of EJCs and peripheral EJC factor RNPS1 to suppress proximal methylation and the ability of EJC and RNPS1 to suppress splicing. We find that EJC-suppressed methylation sites colocalize with EJC-suppressed splice sites, suggesting that EJCs and exon architecture broadly determine accessibility of local mRNA to regulatory machineries through packaging of exon junction-proximal mRNA. Altogether, we identify EJCs as m^{6}A suppressors that control global m^{6}A methylation specificity, revealing important previously unrecognized links between m^{6}A methylation and splicing, exon architecture and mRNP packaging.

Future questions

1. Are there more m^{6}A suppressors, beyond the EJC, that suppress m^{6}A deposition?
2. What determines the length of RNA that EJCs protect from methylation? Are there predictable rules that determine exactly how much RNA is protected?
3. Does EJC-mediated mRNP packaging control general accessibility of RNA to additional regulatory machineries beyond the m^{6}A methylation and splicing machineries?
4. What are the relative contributions of different m^{6}A regulatory mechanisms to the m^{6}A-mediated gene expression dysregulation observed upon EJC depletion?
5. To what extent can variation of mRNA exon architectures and EJC positioning explain differential gene expression outcomes between various transcript isoforms of a single gene or transcripts from different genes?
6. What are the functional effects of differential methylation of canonical transcript isoforms and alternative transcript isoforms with longer exons?
7. Are there other mechanisms, beyond alternative pre-mRNA processing, that enable EJC-suppressed m^{6}A sites to escape EJC suppression?
8. Does hypermethylation of mRNA expressed from expression constructs lacking endogenous splice sites have implications for the use of these constructs in applications such as gene therapy?
9. How might remodeling of EJC composition throughout the lifetimes of mRNPs change accessibility of mRNA to methylation, splicing and possibly other machineries?

REFERENCES

Ae, A., An, D., and Re, K. (2017). RNA Chemical Proteomics Reveals the N6-Methyladenosine (m6A)-Regulated Protein-RNA Interactome. J Am Chem Soc *139*, 17249–17252.

Agarwal, V., Bell, G.W., Nam, J.-W., and Bartel, D.P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. ELife *4*, e05005.

Agarwala, S.D., Blitzblau, H.G., Hochwagen, A., and Fink, G.R. (2012). RNA methylation by the MIS complex regulates a cell fate decision in yeast. PLoS Genet. *8*, e1002732.

Aguilo, F., Zhang, F., Sancho, A., Fidalgo, M., Di Cecilia, S., Vashisht, A., Lee, D.-F., Chen, C.-H., Rengasamy, M., Andino, B., et al. (2015). Coordination of m6A mRNA methylation and gene transcription by ZFP217 regulates pluripotency and reprogramming. Cell Stem Cell *17*, 689–704.

Akhtar, J., Kreim, N., Marini, F., Mohana, G., Brüne, D., Binder, H., and Roignant, J.-Y. (2019). Promoter-proximal pausing mediated by the exon junction complex regulates splicing. Nature Communications *10*, 521.

Alarcón, C.R., Lee, H., Goodarzi, H., Halberg, N., and Tavazoie, S.F. (2015a). N6-methyl-adenosine (m6A) marks primary microRNAs for processing. Nature *519*, 482–485.

Alarcón, C.R., Goodarzi, H., Lee, H., Liu, X., Tavazoie, S., and Tavazoie, S.F. (2015b). HNRNPA2B1 is a mediator of m6A-dependent nuclear RNA processing events. Cell *162*, 1299–1308.

Allis, C.D., and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. Nat Rev Genet *17*, 487–500.

AveXis, Inc. (2020). Phase 3, Open-Label, Single-Arm, Single-Dose Gene Replacement Therapy Clinical Trial for Patients With Spinal Muscular Atrophy Type 1 With One or Two SMN2 Copies Delivering AVXS-101 by Intravenous Infusion (clinicaltrials.gov).

Baltz, A.G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., et al. (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol Cell *46*, 674–690.

Bao, X., Guo, X., Yin, M., Tariq, M., Lai, Y., Kanwal, S., Zhou, J., Li, N., Lv, Y., Pulido-Quetglas, C., et al. (2018). Capturing the interactome of newly transcribed RNA. Nat Methods *15*, 213–220.

Barbieri, I., Tzelepis, K., Pandolfini, L., Shi, J., Millán-Zambrano, G., Robson, S.C., Aspris, D., Migliori, V., Bannister, A.J., Han, N., et al. (2017). Promoter-bound METTL3 maintains myeloid leukaemia by m6A-dependent translation control. Nature *552*, 126–131.

Barnett, S.F., Northington, S.J., and LeStourgeon, W.M. (1990). Isolation and in vitro assembly of nuclear ribonucleoprotein particles and purification of core particle proteins. In Methods in Enzymology, (Academic Press), pp. 293–307.

Bartosovic, M., Molares, H.C., Gregorova, P., Hrossova, D., Kudla, G., and Vanacova, S. (2017). N6-methyladenosine demethylase FTO targets pre-mRNAs and regulates alternative splicing and 3'-end processing. Nucleic Acids Res *45*, 11356–11370.

Batista, P.J., Molinie, B., Wang, J., Qu, K., Zhang, J., Li, L., Bouley, D.M., Lujan, E., Haddad, B., Daneshvar, K., et al. (2014). m6A RNA modification controls cell fate transition in mammalian embryonic stem cells. Cell Stem Cell *15*, 707–719.

Bertero, A., Brown, S., Madrigal, P., Osnato, A., Ortmann, D., Yiangou, L., Kadiwala, J., Hubner, N.C., de Los Mozos, I.R., Sadée, C., et al. (2018a). The SMAD2/3 interactome reveals that TGFβ controls m6A mRNA methylation in pluripotency. Nature *555*, 256–259.

Bertero, A., Brown, S., Madrigal, P., Osnato, A., Ortmann, D., Yiangou, L., Kadiwala, J., Hubner, N.C., de Los Mozos, I.R., Sadée, C., et al. (2018b). The SMAD2/3 interactome reveals that TGFβ controls m6A mRNA methylation in pluripotency. Nature *555*, 256–259.

Beyer, A.L., Christensen, M.E., Walker, B.W., and LeStourgeon, W.M. (1977). Identification and characterization of the packaging proteins of core 40S hnRNP particles. Cell *11*, 127–138.

Bieberstein, N.I., Carrillo Oesterreich, F., Straube, K., and Neugebauer, K.M. (2012). First Exon Length Controls Active Chromatin Signatures and Transcription. Cell Reports *2*, 62–68.

Blazquez, L., Emmett, W., Faraway, R., Pineda, J.M.B., Bajew, S., Gohr, A., Haberman, N., Sibley, C.R., Bradley, R.K., Irimia, M., et al. (2018). Exon Junction Complex Shapes the Transcriptome by Repressing Recursive Splicing. Mol Cell *72*, 496-509.e9.

Boccaletto, P., Machnicka, M.A., Purta, E., Piątkowski, P., Bagiński, B., Wirecki, T.K., de Crécy-Lagard, V., Ross, R., Limbach, P.A., Kotter, A., et al. (2018). MODOMICS: a database of RNA modification pathways. 2017 update. Nucleic Acids Res *46*, D303–D307.

Bodi, Z., Zhong, S., Mehra, S., Song, J., Graham, N., Li, H., May, S., and Fray, R.G. (2012). Adenosine Methylation in Arabidopsis mRNA is Associated with the 3' End and Reduced Levels Cause Developmental Defects. Front Plant Sci *3*, 48.

Boehm, V., and Gehring, N.H. (2016). Exon Junction Complexes: Supervising the Gene Expression Assembly Line. Trends Genet. *32*, 724–735.

Boehm, V., Britto-Borges, T., Steckelberg, A.-L., Singh, K.K., Gerbracht, J.V., Gueney, E., Blazquez, L., Altmüller, J., Dieterich, C., and Gehring, N.H. (2018). Exon Junction Complexes Suppress Spurious Splice Sites to Safeguard Transcriptome Integrity. Molecular Cell *72*, 482-495.e7.

Bokar, J.A., Rath-Shambaugh, M.E., Ludwiczak, R., Narayan, P., and Rottman, F. (1994). Characterization and partial purification of mRNA N6-adenosine methyltransferase from HeLa cell nuclei. Internal mRNA methylation requires a multisubunit complex. J. Biol. Chem. *269*, 17697–17704.

Bokar, J.A., Shambaugh, M.E., Polayes, D., Matera, A.G., and Rottman, F.M. (1997). Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. RNA *3*, 1233–1247.

Bradner, J.E., Hnisz, D., and Young, R.A. (2017). Transcriptional Addiction in Cancer. Cell *168*, 629–643.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nature Biotechnology *34*, 525–527.

Bujnicki, J.M., Feder, M., Radlinska, M., and Blumenthal, R.M. (2002). Structure prediction and phylogenetic analysis of a functionally diverse family of proteins homologous to the MT-A70 subunit of the human mRNA:m(6)A methyltransferase. J. Mol. Evol. *55*, 431–444.

Busch, A., and Hertel, K.J. (2013). HEXEvent: a database of Human EXon splicing Events. Nucleic Acids Res. *41*, D118-124.

Buschmann, T., and Bystrykh, L.V. (2013). Levenshtein error-correcting barcodes for multiplexed DNA sequencing. BMC Bioinformatics *14*, 272.

Carpenter, S., Ricci, E.P., Mercier, B.C., Moore, M.J., and Fitzgerald, K.A. (2014). Post-transcriptional regulation of gene expression in innate immunity. Nat Rev Immunol *14*, 361–376.

Chatterjee, N., and Walker, G.C. (2017). Mechanisms of DNA damage, repair and mutagenesis. Environ Mol Mutagen *58*, 235–263.

Chen, M., Wei, L., Law, C.-T., Tsang, F.H.-C., Shen, J., Cheng, C.L.-H., Tsang, L.-H., Ho, D.W.-H., Chiu, D.K.-C., Lee, J.M.-F., et al. (2018a). RNA N6-methyladenosine methyltransferase-like 3 promotes liver cancer progression through YTHDF2-dependent posttranscriptional silencing of SOCS2. Hepatology *67*, 2254–2270.

Chen, X., Teichmann, S.A., and Meyer, K.B. (2018b). From Tissues to Cell Types and Back: Single-Cell Gene Expression Analysis of Tissue Architecture. Annual Review of Biomedical Data Science *1*, 29–51.

Cheng, Y., Luo, H., Izzo, F., Pickering, B.F., Nguyen, D., Myers, R., Schurer, A., Gourkanti, S., Brüning, J.C., Vu, L.P., et al. (2019). m6A RNA Methylation Maintains Hematopoietic Stem Cell Identity and Symmetric Commitment. Cell Rep *28*, 1703-1716.e6.

Conway, G., Wooley, J., Bibring, T., and LeStourgeon, W.M. (1988). Ribonucleoproteins package 700 nucleotides of pre-mRNA into a repeating array of regular particles. Mol Cell Biol *8*, 2884–2895.

Cui, Q., Shi, H., Ye, P., Li, L., Qu, Q., Sun, G., Sun, G., Lu, Z., Huang, Y., Yang, C.-G., et al. (2017). m6A RNA Methylation Regulates the Self-Renewal and Tumorigenesis of Glioblastoma Stem Cells. Cell Reports *18*, 2622–2634.

Cully, M. (2019). Chemical inhibitors make their RNA epigenetic mark. Nature Reviews Drug Discovery *18*, 892–894.

Desrosiers, R., Friderici, K., and Rottman, F. (1974). Identification of Methylated Nucleosides in Messenger RNA from Novikoff Hepatoma Cells. PNAS *71*, 3971–3975.

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature *485*, 201–206.

Dorn, L.E., Lasman, L., Chen, J., Xu, X., Hund, T.J., Medvedovic, M., Hanna, J.H., van Berlo, J.H., and Accornero, F. (2019). The N6-Methyladenosine mRNA Methylase METTL3 Controls Cardiac Homeostasis and Hypertrophy. Circulation *139*, 533–545.

Dostie, J., and Dreyfuss, G. (2002). Translation is required to remove Y14 from mRNAs in the cytoplasm. Curr Biol *12*, 1060–1067.

Du, H., Zhao, Y., He, J., Zhang, Y., Xi, H., Liu, M., Ma, J., and Wu, L. (2016). YTHDF2 destabilizes m^6A-containing RNA through direct recruitment of the CCR4–NOT deadenylase complex. Nature Communications *7*, 12626.

Du, Y., Hou, G., Zhang, H., Dou, J., He, J., Guo, Y., Li, L., Chen, R., Wang, Y., Deng, R., et al. (2018). SUMOylation of the m6A-RNA methyltransferase METTL3 modulates its function. Nucleic Acids Res *46*, 5195–5208.

Edupuganti, R.R., Geiger, S., Lindeboom, R.G.H., Shi, H., Hsu, P.J., Lu, Z., Wang, S.-Y., Baltissen, M.P.A., Jansen, P.W.T.C., Rossa, M., et al. (2017). N6-methyladenosine (m6A) recruits and repels proteins to regulate mRNA homeostasis. Nat. Struct. Mol. Biol. *24*, 870–878.

Engel, M., Eggert, C., Kaplick, P.M., Eder, M., Röh, S., Tietze, L., Namendorf, C., Arloth, J., Weber, P., Rex-Haffner, M., et al. (2018). The Role of m6A/m-RNA Methylation in Stress Response Regulation. Neuron *99*, 389-403.e9.

Fish, L., Navickas, A., Culbertson, B., Xu, Y., Nguyen, H.C.B., Zhang, S., Hochman, M., Okimoto, R., Dill, B.D., Molina, H., et al. (2019). Nuclear TARBP2 Drives Oncogenic Dysregulation of RNA Splicing and Decay. Mol. Cell *75*, 967-981.e9.

Frye, M., Harada, B.T., Behm, M., and He, C. (2018). RNA modifications modulate gene expression during development. Science *361*, 1346–1349.

Garcia-Campos, M.A., Edelheit, S., Toth, U., Safra, M., Shachar, R., Viukov, S., Winkler, R., Nir, R., Lasman, L., Brandis, A., et al. (2019). Deciphering the “m6A Code” via Antibody-Independent Quantitative Profiling. Cell *178*, 731-747.e16.

Gehring, N.H., Wahle, E., and Fischer, U. (2017). Deciphering the mRNP Code: RNA-Bound Determinants of Post-Transcriptional Gene Regulation. Trends in Biochemical Sciences *42*, 369–382.

Geula, S., Moshitch-Moshkovitz, S., Dominissini, D., Mansour, A.A., Kol, N., Salmon-Divon, M., Hershkovitz, V., Peer, E., Mor, N., Manor, Y.S., et al. (2015). m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. Science *347*, 1002–1006.

Gilbert, W.V., Bell, T.A., and Schaening, C. (2016). Messenger RNA modifications: Form, distribution, and function. Science *352*, 1408–1412.

Guo, J., Tang, H.-W., Li, J., Perrimon, N., and Yan, D. (2018). Xio is a component of the Drosophila sex determination pathway and RNA N6-methyladenosine methyltransferase complex. Proc. Natl. Acad. Sci. U.S.A. *115*, 3674–3679.

He, P.C., and He, C. (2021). m6A RNA methylation: from mechanisms to therapeutic potential. The EMBO Journal *n/a*, e105977.

Herrmann, C.J., Schmidt, R., Kanitz, A., Artimo, P., Gruber, A.J., and Zavolan, M. (2020). PolyASite 2.0: a consolidated atlas of polyadenylation sites from 3′ end sequencing. Nucleic Acids Research *48*, D174–D179.

Hesser, C.R., Karijolich, J., Dominissini, D., He, C., and Glaunsinger, B.A. (2018). N6-methyladenosine modification and the YTHDF2 reader protein play cell type specific roles in lytic viral gene expression during Kaposi's sarcoma-associated herpesvirus infection. PLOS Pathogens *14*, e1006995.

Hir, H.L., Saulière, J., and Wang, Z. (2016). The exon junction complex as a node of post-transcriptional networks. Nature Reviews Molecular Cell Biology *17*, 41–54.

Hongay, C.F., and Orr-Weaver, T.L. (2011). Drosophila Inducer of MEiosis 4 (IME4) is required for Notch signaling during oogenesis. Proc. Natl. Acad. Sci. U.S.A. *108*, 14855–14860.

Hsu, P.J., Zhu, Y., Ma, H., Guo, Y., Shi, X., Liu, Y., Qi, M., Lu, Z., Shi, H., Wang, J., et al. (2017). Ythdc2 is an N6-methyladenosine binding protein that regulates mammalian spermatogenesis. Cell Res. *27*, 1115–1127.

Huang, H., Weng, H., Zhou, K., Wu, T., Zhao, B.S., Sun, M., Chen, Z., Deng, X., Xiao, G., Auer, F., et al. (2019). Histone H3 trimethylation at lysine 36 guides m6A RNA modification co-transcriptionally. Nature *567*, 414–419.

Jackson, R.J., Hellen, C.U.T., and Pestova, T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. Nat Rev Mol Cell Biol *11*, 113–127.

Kasowitz, S.D., Ma, J., Anderson, S.J., Leu, N.A., Xu, Y., Gregory, B.D., Schultz, R.M., and Wang, P.J. (2018). Nuclear m6A reader YTHDC1 regulates alternative polyadenylation and splicing during mouse oocyte development. PLOS Genetics *14*, e1007412.

Ke, S., Alemu, E.A., Mertens, C., Gantman, E.C., Fak, J.J., Mele, A., Haripal, B., Zucker-Scharff, I., Moore, M.J., Park, C.Y., et al. (2015). A majority of m6A residues are in the last exons, allowing the potential for 3′ UTR regulation. Genes Dev *29*, 2037–2053.

Ke, S., Pandya-Jones, A., Saito, Y., Fak, J.J., Vågbø, C.B., Geula, S., Hanna, J.H., Black, D.L., Darnell, J.E., and Darnell, R.B. (2017). m6A mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover. Genes Dev. *31*, 990–1006.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, and D. (2002). The Human Genome Browser at UCSC. Genome Res. *12*, 996–1006.

Kierzek, E., and Kierzek, R. (2003). The thermodynamic stability of RNA duplexes and hairpins containing N6-alkyladenosines and 2-methylthio-N6-alkyladenosines. Nucleic Acids Res. *31*, 4472–4480.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nature Methods *12*, 357–360.

Knuckles, P., Carl, S.H., Musheev, M., Niehrs, C., Wenger, A., and Bühler, M. (2017). RNA fate determination through cotranscriptional adenosine methylation and microprocessor binding. Nature Structural & Molecular Biology *24*, 561–569.

Knuckles, P., Lence, T., Haussmann, I.U., Jacob, D., Kreim, N., Carl, S.H., Masiello, I., Hares, T., Villaseñor, R., Hess, D., et al. (2018). Zc3h13/Flacc is required for adenosine methylation by bridging the mRNA-binding factor Rbm15/Spenito to the m6A machinery component Wtap/Fl(2)d. Genes Dev. *32*, 415–429.

König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M., and Ule, J. (2010). iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nat Struct Mol Biol *17*, 909–915.

Kornberg, R.D., and Lorch, Y. (1999). Twenty-Five Years of the Nucleosome, Fundamental Particle of the Eukaryote Chromosome. Cell *98*, 285–294.

Kretschmer, J., Rao, H., Hackert, P., Sloan, K.E., Höbartner, C., and Bohnsack, M.T. (2018). The m6A reader protein YTHDC2 interacts with the small ribosomal subunit and the 5′–3′ exoribonuclease XRN1. RNA *24*, 1339–1350.

Lavi, S., and Shatkin, A.J. (1975). Methylated simian virus 40-specific RNA from nuclei and cytoplasm of infected BSC-1 cells. Proc. Natl. Acad. Sci. U.S.A. *72*, 2012–2016.

Le Hir, H., Izaurralde, E., Maquat, L.E., and Moore, M.J. (2000). The spliceosome deposits multiple proteins 20–24 nucleotides upstream of mRNA exon–exon junctions. The EMBO Journal *19*, 6860–6869.

Ledford, H. (2018). First test of in-body gene editing shows promise. Nature.

Lee, T.I., and Young, R.A. (2013). Transcriptional Regulation and Its Misregulation in Disease. Cell *152*, 1237–1251.

Lee, H., Bao, S., Qian, Y., Geula, S., Leslie, J., Zhang, C., Hanna, J.H., and Ding, L. (2019). Stage-specific requirement for Mettl3 -dependent m 6 A mRNA methylation during haematopoietic stem cell differentiation. Nature Cell Biology *21*, 700–709.

Lee, S.-H., Singh, I., Tisdale, S., Abdel-Wahab, O., Leslie, C.S., and Mayr, C. (2018). Widespread intronic polyadenylation inactivates tumour suppressor genes in leukaemia. Nature *561*, 127–131.

Lee, W.-C., Hou, B.-H., Hou, C.-Y., Tsao, S.-M., Kao, P., and Chen, H.-M. (2020). Widespread Exon Junction Complex Footprints in the RNA Degradome Mark mRNA Degradation before Steady State Translation. The Plant Cell *32*, 904–922.

Lejeune, F., Ishigaki, Y., Li, X., and Maquat, L.E. (2002). The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: dynamics of mRNP remodeling. EMBO J *21*, 3536–3545.

Li, H.-B., Tong, J., Zhu, S., Batista, P.J., Duffy, E.E., Zhao, J., Bailis, W., Cao, G., Kroehling, L., Chen, Y., et al. (2017a). m6A mRNA methylation controls T cell homeostasis by targeting the IL-7/STAT5/SOCS pathways. Nature *548*, 338–342.

Li, X., Xiong, X., and Yi, C. (2017b). Epitranscriptome sequencing technologies: decoding RNA modifications. Nat Methods *14*, 23–31.

Li, X., Tang, J., Huang, W., Wang, F., Li, P., Qin, C., Qin, Z., Zou, Q., Wei, J., Hua, L., et al. (2017c). The M6A methyltransferase METTL3: acting as a tumor suppressor in renal cell carcinoma. Oncotarget *8*, 96103–96116.

Lichinchi, G., Gao, S., Saletore, Y., Gonzalez, G.M., Bansal, V., Wang, Y., Mason, C.E., and Rana, T.M. (2016a). Dynamics of the human and viral m 6 A RNA methylomes during HIV-1 infection of T cells. Nat Microbiol *1*, 1–9.

Lichinchi, G., Zhao, B.S., Wu, Y., Lu, Z., Qin, Y., He, C., and Rana, T.M. (2016b). Dynamics of Human and Viral RNA Methylation during Zika Virus Infection. Cell Host & Microbe *20*, 666–673.

Lin, Z., Hsu, P.J., Xing, X., Fang, J., Lu, Z., Zou, Q., Zhang, K.-J., Zhang, X., Zhou, Y., Zhang, T., et al. (2017). Mettl3-/Mettl14-mediated mRNA N 6 -methyladenosine modulates murine spermatogenesis. Cell Research *27*, 1216–1230.

Linder, B., Grozhik, A.V., Olarerin-George, A.O., Meydan, C., Mason, C.E., and Jaffrey, S.R. (2015). Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. Nat. Methods *12*, 767–772.

Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., Jia, G., Yu, M., Lu, Z., Deng, X., et al. (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat Chem Biol *10*, 93–95.

Liu, J., Eckert, M.A., Harada, B.T., Liu, S.-M., Lu, Z., Yu, K., Tienda, S.M., Chryplewicz, A., Zhu, A.C., Yang, Y., et al. (2018). m6A mRNA methylation regulates AKT activity to promote the proliferation and tumorigenicity of endometrial cancer. Nat. Cell Biol. *20*, 1074–1083.

Liu, J., Dou, X., Chen, C., Chen, C., Liu, C., Xu, M.M., Zhao, S., Shen, B., Gao, Y., Han, D., et al. (2020a). N6-methyladenosine of chromosome-associated regulatory RNA regulates chromatin state and transcription. Science *367*, 580–586.

Liu, J., Li, K., Cai, J., Zhang, M., Zhang, X., Xiong, X., Meng, H., Xu, X., Huang, Z., Peng, J., et al. (2020b). Landscape and Regulation of m6A and m6Am Methylome across Human and Mouse Tissues. Molecular Cell *77*, 426-440.e6.

Liu, N., Parisien, M., Dai, Q., Zheng, G., He, C., and Pan, T. (2013). Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. RNA *19*, 1848–1856.

Liu, N., Dai, Q., Zheng, G., He, C., Parisien, M., and Pan, T. (2015). N6-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. Nature *518*, 560–564.

Liu, N., Zhou, K.I., Parisien, M., Dai, Q., Diatchenko, L., and Pan, T. (2017). N6-methyladenosine alters RNA structure to regulate binding of a low-complexity protein. Nucleic Acids Res. *45*, 6051–6063.

Louloupi, A., Ntini, E., Conrad, T., and Ørom, U.A.V. (2018). Transient N-6-Methyladenosine Transcriptome Sequencing Reveals a Regulatory Role of m6A in Splicing Efficiency. Cell Rep *23*, 3429–3437.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Lu, C.-C., Lee, C.-C., Tseng, C.-T., and Tarn, W.-Y. (2017). Y14 governs p53 expression and modulates DNA damage sensitivity. Scientific Reports *7*, 45558.

Ma, J.-Z., Yang, F., Zhou, C.-C., Liu, F., Yuan, J.-H., Wang, F., Wang, T.-T., Xu, Q.-G., Zhou, W.-P., and Sun, S.-H. (2017). METTL14 suppresses the metastatic potential of hepatocellular carcinoma by modulating N6 -methyladenosine-dependent primary MicroRNA processing. Hepatology *65*, 529–543.

Mabin, J.W., Woodward, L.A., Patton, R.D., Yi, Z., Jia, M., Wysocki, V.H., Bundschuh, R., and Singh, G. (2018). The Exon Junction Complex Undergoes a Compositional Switch that Alters mRNP Structure and Nonsense-Mediated mRNA Decay Activity. Cell Rep *25*, 2431-2446.e7.

Malcolm, D.B., and Sommerville, J. (1974). The structure of chromosome-derived ribonucleoprotein in oocytes of Triturus cristatus carnifex (Laurenti). Chromosoma *48*, 137–158.

Mao, H., McMahon, J.J., Tsai, Y.-H., Wang, Z., and Silver, D.L. (2016). Haploinsufficiency for Core Exon Junction Complex Components Disrupts Embryonic Neurogenesis and Causes p53-Mediated Microcephaly. PLoS Genet. *12*, e1006282.

Mao, Y., Dong, L., Liu, X.-M., Guo, J., Ma, H., Shen, B., and Qian, S.-B. (2019). m 6 A in mRNA coding regions promotes translation via the RNA helicase-containing YTHDC2. Nature Communications *10*, 1–11.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal *17*, 10–12.

Mauer, J., Luo, X., Blanjoie, A., Jiao, X., Grozhik, A.V., Patil, D.P., Linder, B., Pickering, B.F., Vasseur, J.-J., Chen, Q., et al. (2017). Reversible methylation of m6Am in the 5' cap controls mRNA stability. Nature *541*, 371–375.

Mauer, J., Sindelar, M., Despic, V., Guez, T., Hawley, B.R., Vasseur, J.-J., Rentmeister, A., Gross, S.S., Pellizzoni, L., Debart, F., et al. (2019). FTO controls reversible m6Am RNA methylation during snRNA biogenesis. Nat. Chem. Biol. *15*, 340–347.

McAfee, J.G., Shahied-Milam, L., Soltaninassab, S.R., and LeStourgeon, W.M. (1996). A major determinant of hnRNP C protein binding to RNA is a novel bZIP-like RNA binding domain. RNA *2*, 1139–1152.

McIntyre, A.B.R., Gokhale, N.S., Cerchietti, L., Jaffrey, S.R., Horner, S.M., and Mason, C.E. (2020). Limits in the detection of m 6 A changes using MeRIP/m 6 A-seq. Sci Rep *10*, 6590.

McKnight, S.L., and Miller, O.L. (1976). Ultrastructural patterns of RNA synthesis during early embryogenesis of Drosophila melanogaster. Cell *8*, 305–319.

McMahon, J.J., Miller, E.E., and Silver, D.L. (2016). The exon junction complex in neural development and neurodevelopmental disease. International Journal of Developmental Neuroscience *55*, 117–123.

Mercuri, E., Pera, M.C., Scoto, M., Finkel, R., and Muntoni, F. (2020). Spinal muscular atrophy — insights and challenges in the treatment era. Nature Reviews Neurology *16*, 706–715.

Metkar, M., Ozadam, H., Lajoie, B.R., Imakaev, M., Mirny, L.A., Dekker, J., and Moore, M.J. (2018). Higher-Order Organization Principles of Pre-translational mRNPs. Mol Cell *72*, 715-726.e3.

Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E., and Jaffrey, S.R. (2012). Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3' UTRs and Near Stop Codons. Cell *149*, 1635–1646.

Meyer, K.D., Patil, D.P., Zhou, J., Zinoviev, A., Skabkin, M.A., Elemento, O., Pestova, T.V., Qian, S.-B., and Jaffrey, S.R. (2015). 5' UTR m(6)A Promotes Cap-Independent Translation. Cell *163*, 999–1010.

Middleton, R., Gao, D., Thomas, A., Singh, B., Au, A., Wong, J.J.-L., Bomane, A., Cosson, B., Eyras, E., Rasko, J.E.J., et al. (2017). IRFinder: assessing the impact of intron retention on mammalian gene expression. Genome Biology *18*, 51.

Molinie, B., Wang, J., Lim, K.S., Hillebrand, R., Lu, Z., Van Wittenberghe, N., Howard, B.D., Daneshvar, K., Mullen, A.C., Dedon, P., et al. (2016). m 6 A-LAIC-seq reveals the census and complexity of the m 6 A epitranscriptome. Nature Methods *13*, 692–698.

Monteuuis, G., Wong, J.J.L., Bailey, C.G., Schmitz, U., and Rasko, J.E.J. (2019). The changing paradigm of intron retention: regulation, ramifications and recipes. Nucleic Acids Research *47*, 11497–11513.

Nance, K.D., and Meier, J.L. (2021). Modifications in an Emergency: The Role of N1-Methylpseudouridine in COVID-19 Vaccines. ACS Cent Sci *7*, 748–756.

Naro, C., Jolly, A., Di Persio, S., Bielli, P., Setterblad, N., Alberdi, A.J., Vicini, E., Geremia, R., De la Grange, P., and Sette, C. (2017). An Orchestrated Intron Retention Program in Meiosis Controls Timely Usage of Transcripts during Germ Cell Differentiation. Developmental Cell *41*, 82-93.e4.

Ontiveros, R.J., Stoute, J., and Liu, K.F. (2019). The chemical diversity of RNA modifications. Biochemical Journal *476*, 1227–1245.

Park, O.H., Ha, H., Lee, Y., Boo, S.H., Kwon, D.H., Song, H.K., and Kim, Y.K. (2019). Endoribonucleolytic Cleavage of m6A-Containing RNAs by RNase P/MRP Complex. Mol. Cell *74*, 494-507.e8.

Patil, D.P., Chen, C.-K., Pickering, B.F., Chow, A., Jackson, C., Guttman, M., and Jaffrey, S.R. (2016). m(6)A RNA methylation promotes XIST-mediated transcriptional repression. Nature *537*, 369–373.

Patton, R.D., Sanjeev, M., Woodward, L.A., Mabin, J.W., Bundschuh, R., and Singh, G. (2020). Chemical crosslinking enhances RNA immunoprecipitation for efficient identification of binding sites of proteins that photo-crosslink poorly with RNA. RNA rna.074856.120.

Perry, R.P., and Kelley, D.E. (1966). Buoyant densities of cytoplasmic ribonucleoprotein particles of mammalian cells: Distinctive character of ribosome subunits and the rapidly labeled components. Journal of Molecular Biology *16*, 255–268.

Perry, R.P., Kelley, D.E., Friderici, K., and Rottman, F. (1975). The methylated constituents of L cell messenger RNA: Evidence for an unusual cluster at the 5′ terminus. Cell *4*, 387–394.

Prabakaran, S., Lippens, G., Steen, H., and Gunawardena, J. (2012). Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. Wiley Interdiscip Rev Syst Biol Med *4*, 565–583.

Queiroz, R.M.L., Smith, T., Villanueva, E., Marti-Solano, M., Monti, M., Pizzinga, M., Mirea, D.-M., Ramakrishna, M., Harvey, R.F., Dezi, V., et al. (2019). Comprehensive identification of RNA–protein interactions in any organism using orthogonal organic phase separation (OOPS). Nat Biotechnol *37*, 169–178.

Ramsköld, D., Wang, E.T., Burge, C.B., and Sandberg, R. (2009). An Abundance of Ubiquitously Expressed Genes Revealed by Tissue Transcriptome Sequence Data. PLoS Comput Biol *5*.

Ries, R.J., Zaccara, S., Klein, P., Olarerin-George, A., Namkoong, S., Pickering, B.F., Patil, D.P., Kwak, H., Lee, J.H., and Jaffrey, S.R. (2019). m 6 A enhances the phase separation potential of mRNA. Nature *571*, 424–428.

Roignant, J.-Y., and Treisman, J.E. (2010). Exon Junction Complex Subunits Are Required to Splice Drosophila MAP Kinase, a Large Heterochromatic Gene. Cell *143*, 238–250.

Roost, C., Lynch, S.R., Batista, P.J., Qu, K., Chang, H.Y., and Kool, E.T. (2015). Structure and thermodynamics of N6-methyladenosine in RNA: a spring-loaded base modification. J. Am. Chem. Soc. *137*, 2107–2115.

Roundtree, I.A., Evans, M.E., Pan, T., and He, C. (2017a). Dynamic RNA Modifications in Gene Expression Regulation. Cell *169*, 1187–1200.

Roundtree, I.A., Luo, G.-Z., Zhang, Z., Wang, X., Zhou, T., Cui, Y., Sha, J., Huang, X., Guerrero, L., Xie, P., et al. (2017b). YTHDC1 mediates nuclear export of N6-methyladenosine methylated mRNAs. Elife *6*.

Rubio, R.M., Depledge, D.P., Bianco, C., Thompson, L., and Mohr, I. (2018). RNA m6 A modification enzymes shape innate responses to DNA by regulating interferon β. Genes Dev. *32*, 1472–1484.

Růžička, K., Zhang, M., Campilho, A., Bodi, Z., Kashif, M., Saleh, M., Eeckhout, D., El-Showk, S., Li, H., Zhong, S., et al. (2017). Identification of factors required for m6 A mRNA methylation in Arabidopsis reveals a role for the conserved E3 ubiquitin ligase HAKAI. New Phytol. *215*, 157–172.

Saulière, J., Murigneux, V., Wang, Z., Marquenet, E., Barbosa, I., Le Tonquèze, O., Audic, Y., Paillard, L., Roest Crollius, H., and Le Hir, H. (2012). CLIP-seq of eIF4AIII reveals transcriptome-wide mapping of the human exon junction complex. Nat Struct Mol Biol *19*, 1124–1131.

Scarpa, M. (1993). Mucopolysaccharidosis Type II. In GeneReviews®, M.P. Adam, H.H. Ardinger, R.A. Pagon, S.E. Wallace, L.J. Bean, G. Mirzaa, and A. Amemiya, eds. (Seattle (WA): University of Washington, Seattle), p.

Schlautmann, L.P., and Gehring, N.H. (2020). A Day in the Life of the Exon Junction Complex. Biomolecules *10*, 866.

Schwartz, S., Meshorer, E., and Ast, G. (2009). Chromatin organization marks exon-intron structure. Nat Struct Mol Biol *16*, 990–995.

Schwartz, S., Mumbach, M.R., Jovanovic, M., Wang, T., Maciag, K., Bushkin, G.G., Mertins, P., Ter-Ovanesyan, D., Habib, N., Cacchiarelli, D., et al. (2014). Perturbation of m6A Writers Reveals Two Distinct Classes of mRNA Methylation at Internal and 5′ Sites. Cell Reports *8*, 284–296.

Shi, H., Wang, X., Lu, Z., Zhao, B.S., Ma, H., Hsu, P.J., Liu, C., and He, C. (2017). YTHDF3 facilitates translation and decay of N6-methyladenosine-modified RNA. Cell Res. *27*, 315–328.

Shi, H., Wei, J., and He, C. (2019). Where, When, and How: Context-Dependent Functions of RNA Methylation Writers, Readers, and Erasers. Molecular Cell *74*, 640–650.

Silver, D.L., Watkins-Chow, D.E., Schreck, K.C., Pierfelice, T.J., Larson, D.M., Burnetti, A.J., Liaw, H.-J., Myung, K., Walsh, C.A., Gaiano, N., et al. (2010). The exon junction complex component Magoh controls brain size by regulating neural stem cell division. Nat Neurosci *13*, 551–558.

Singh, G., Kucukural, A., Cenik, C., Leszyk, J.D., Shaffer, S.A., Weng, Z., and Moore, M.J. (2012). The Cellular EJC Interactome Reveals Higher Order mRNP Structure and an EJC-SR Protein Nexus. Cell *151*, 750–764.

Singh, G., Pratt, G., Yeo, G.W., and Moore, M.J. (2015). The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. Annual Review of Biochemistry *84*, 325–354.

Singh, I., Lee, S.-H., Sperling, A.S., Samur, M.K., Tai, Y.-T., Fulciniti, M., Munshi, N.C., Mayr, C., and Leslie, C.S. (2018). Widespread intronic polyadenylation diversifies immune cell transcriptomes. Nat Commun *9*.

Slobodin, B., Han, R., Calderone, V., Vrielink, J.A.F.O., Loayza-Puch, F., Elkon, R., and Agami, R. (2017a). Transcription Impacts the Efficiency of mRNA Translation via Co-transcriptional N6-adenosine Methylation. Cell *169*, 326-337.e12.

Slobodin, B., Han, R., Calderone, V., Vrielink, J.A.F.O., Loayza-Puch, F., Elkon, R., and Agami, R. (2017b). Transcription Impacts the Efficiency of mRNA Translation via Co-transcriptional N6-adenosine Methylation. Cell *169*, 326-337.e12.

Spirin, A.S. (1979). Messenger ribonucleoproteins (informosomes) and RNA-binding proteins. Mol Biol Rep *5*, 53–57.

Spitale, R.C., Flynn, R.A., Zhang, Q.C., Crisalli, P., Lee, B., Jung, J.-W., Kuchelmeister, H.Y., Batista, P.J., Torre, E.A., Kool, E.T., et al. (2015). Structural imprints in vivo decode RNA regulatory mechanisms. Nature *519*, 486–490.

Sun, L., Fazal, F.M., Li, P., Broughton, J.P., Lee, B., Tang, L., Huang, W., Kool, E.T., Chang, H.Y., and Zhang, Q.C. (2019). RNA structure maps across mammalian cellular compartments. Nat. Struct. Mol. Biol. *26*, 322–330.

Suzuki, M.M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet *9*, 465–476.

Tan, B., Liu, H., Zhang, S., da Silva, S.R., Zhang, L., Meng, J., Cui, X., Yuan, H., Sorel, O., Zhang, S.-W., et al. (2018). Viral and cellular N 6 -methyladenosine and N 6 ,2′- O -dimethyladenosine epitranscriptomes in the KSHV life cycle. Nat Microbiol *3*, 108–120.

Tang, C., Klukovich, R., Peng, H., Wang, Z., Yu, T., Zhang, Y., Zheng, H., Klungland, A., and Yan, W. (2018). ALKBH5-dependent m6A demethylation controls splicing and stability of long 3′-UTR mRNAs in male germ cells. PNAS *115*, E325–E333.

Tange, T.Ø., Shibuya, T., Jurica, M.S., and Moore, M.J. (2005). Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. RNA *11*, 1869–1883.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75–82.

Tilgner, H., Nikolaou, C., Althammer, S., Sammeth, M., Beato, M., Valcárcel, J., and Guigó, R. (2009). Nucleosome positioning as a determinant of exon recognition. Nature Structural & Molecular Biology *16*, 996–1001.

Trendel, J., Schwarzl, T., Horos, R., Prakash, A., Bateman, A., Hentze, M.W., and Krijgsveld, J. (2019). The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. Cell *176*, 391-403.e19.

Uphoff, C.C., and Drexler, H.G. (2004). Detecting Mycoplasma contamination in cell cultures by polymerase chain reaction. Methods Mol. Med. *88*, 319–326.

Urdaneta, E.C., Vieira-Vieira, C.H., Hick, T., Wessels, H.-H., Figini, D., Moschall, R., Medenbach, J., Ohler, U., Granneman, S., Selbach, M., et al. (2019). Purification of cross-linked RNA-protein complexes by phenol-toluol extraction. Nat Commun *10*, 990.

Van Nostrand, E.L., Freese, P., Pratt, G.A., Wang, X., Wei, X., Xiao, R., Blue, S.M., Chen, J.-Y., Cody, N.A.L., Dominguez, D., et al. (2020). A large-scale binding and functional map of human RNA-binding proteins. Nature *583*, 711–719.

Vu, L.P., Pickering, B.F., Cheng, Y., Zaccara, S., Nguyen, D., Minuesa, G., Chou, T., Chow, A., Saletore, Y., MacKay, M., et al. (2017). The N6-methyladenosine (m6A)-forming enzyme

METTL3 controls myeloid differentiation of normal hematopoietic and leukemia cells. Nat. Med. *23*, 1369–1376.

Wang, C.-X., Cui, G.-S., Liu, X., Xu, K., Wang, M., Zhang, X.-X., Jiang, L.-Y., Li, A., Yang, Y., Lai, W.-Y., et al. (2018a). METTL3-mediated m6A modification is required for cerebellar development. PLOS Biology *16*, e2004880.

Wang, H., Hu, X., Huang, M., Liu, J., Gu, Y., Ma, L., Zhou, Q., and Cao, X. (2019). Mettl3-mediated mRNA m 6 A methylation promotes dendritic cell activation. Nature Communications *10*, 1–12.

Wang, P., Doxtader, K.A., and Nam, Y. (2016a). Structural Basis for Cooperative Function of Mettl3 and Mettl14 Methyltransferases. Mol. Cell *63*, 306–317.

Wang, X., Lu, Z., Gomez, A., Hon, G.C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G., et al. (2014a). N6-methyladenosine-dependent regulation of messenger RNA stability. Nature *505*, 117–120.

Wang, X., Zhao, B.S., Roundtree, I.A., Lu, Z., Han, D., Ma, H., Weng, X., Chen, K., Shi, H., and He, C. (2015). N6-methyladenosine Modulates Messenger RNA Translation Efficiency. Cell *161*, 1388–1399.

Wang, X., Feng, J., Xue, Y., Guan, Z., Zhang, D., Liu, Z., Gong, Z., Wang, Q., Huang, J., Tang, C., et al. (2016b). Structural basis of N(6)-adenosine methylation by the METTL3-METTL14 complex. Nature *534*, 575–578.

Wang, Y., Li, Y., Toth, J.I., Petroski, M.D., Zhang, Z., and Zhao, J.C. (2014b). N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. Nat. Cell Biol. *16*, 191–198.

Wang, Y., Li, Y., Yue, M., Wang, J., Kumar, S., Wechsler-Reya, R.J., Zhang, Z., Ogawa, Y., Kellis, M., Duester, G., et al. (2018b). N6-methyladenosine RNA modification regulates embryonic neural stem cell self-renewal through histone modifications. Nat. Neurosci. *21*, 195–206.

Wang, Z., Murigneux, V., and Le Hir, H. (2014c). Transcriptome-wide modulation of splicing by the exon junction complex. Genome Biol. *15*, 551.

Wei, C.M., and Moss, B. (1977). Nucleotide sequences at the N6-methyladenosine sites of HeLa cell messenger ribonucleic acid. Biochemistry *16*, 1672–1676.

Wei, C.M., Gershowitz, A., and Moss, B. (1975). Methylated nucleotides block 5’ terminus of HeLa cell messenger RNA. Cell *4*, 379–386.

Wei, J., Liu, F., Lu, Z., Fei, Q., Ai, Y., He, P.C., Shi, H., Cui, X., Su, R., Klungland, A., et al. (2018). Differential m6A, m6Am, and m1A Demethylation Mediated by FTO in the Cell Nucleus and Cytoplasm. Mol. Cell *71*, 973-985.e5.

Weighardt, F., Biamonti, G., and Riva, S. (1996). The roles of heterogeneous nuclear ribonucleoproteins (hnRNP) in RNA metabolism. BioEssays *18*, 747–756.

Wen, J., Lv, R., Ma, H., Shen, H., He, C., Wang, J., Jiao, F., Liu, H., Yang, P., Tan, L., et al. (2018). Zc3h13 Regulates Nuclear RNA m6A Methylation and Mouse Embryonic Stem Cell Self-Renewal. Mol. Cell *69*, 1028-1038.e6.

Weng, H., Huang, H., Wu, H., Qin, X., Zhao, B.S., Dong, L., Shi, H., Skibbe, J., Shen, C., Hu, C., et al. (2018). METTL14 Inhibits Hematopoietic Stem/Progenitor Differentiation and Promotes Leukemogenesis via mRNA m6A Modification. Cell Stem Cell *22*, 191-205.e9.

Winkler, R., Gillis, E., Lasman, L., Safra, M., Geula, S., Soyris, C., Nachshon, A., Tai-Schmiedel, J., Friedman, N., Le-Trilling, V.T.K., et al. (2019). m6A modification controls the innate immune response to infection by targeting type I interferons. Nat. Immunol. *20*, 173–182.

Wojtas, M.N., Pandey, R.R., Mendel, M., Homolka, D., Sachidanandam, R., and Pillai, R.S. (2017). Regulation of m6A Transcripts by the 3′→5′ RNA Helicase YTHDC2 Is Essential for a Successful Meiotic Program in the Mammalian Germline. Molecular Cell *68*, 374-387.e12.

Wong, J.J.-L., Ritchie, W., Ebner, O.A., Selbach, M., Wong, J.W.H., Huang, Y., Gao, D., Pinello, N., Gonzalez, M., Baidya, K., et al. (2013). Orchestrated Intron Retention Regulates Normal Granulocyte Differentiation. Cell *154*, 583–595.

Woodward, L.A., Mabin, J.W., Gangras, P., and Singh, G. (2017). The exon junction complex: a lifelong guardian of mRNA fate. Wiley Interdiscip Rev RNA *8*.

Wu, B., Su, S., Patil, D.P., Liu, H., Gan, J., Jaffrey, S.R., and Ma, J. (2018a). Molecular basis for the specific and multivariant recognitions of RNA substrates by human hnRNP A2/B1. Nature Communications *9*, 1–12.

Wu, Y., Xie, L., Wang, M., Xiong, Q., Guo, Y., Liang, Y., Li, J., Sheng, R., Deng, P., Wang, Y., et al. (2018b). Mettl3-mediated m 6 A RNA methylation regulates the fate of bone marrow mesenchymal stem cells and osteoporosis. Nature Communications *9*, 1–12.

Xiao, S., Cao, S., Huang, Q., Xia, L., Deng, M., Yang, M., Jia, G., Liu, X., Shi, J., Wang, W., et al. (2019). The RNA N6-methyladenosine modification landscape of human fetal tissues. Nat. Cell Biol. *21*, 651–661.

Xiao, W., Adhikari, S., Dahal, U., Chen, Y.-S., Hao, Y.-J., Sun, B.-F., Sun, H.-Y., Li, A., Ping, X.-L., Lai, W.-Y., et al. (2016). Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing. Mol. Cell *61*, 507–519.

Xiao, Y., Wang, Y., Tang, Q., Wei, L., Zhang, X., and Jia, G. (2018). An Elongation- and Ligation-Based qPCR Amplification Method for the Radiolabeling-Free Detection of Locus-Specific N6-Methyladenosine Modification. Angewandte Chemie International Edition *57*, 15995–16000.

Xu, H., Dzhashiashvili, Y., Shah, A., Kunjamma, R.B., Weng, Y., Elbaz, B., Fei, Q., Jones, J.S., Li, Y.I., Zhuang, X., et al. (2020). m6A mRNA Methylation Is Essential for Oligodendrocyte Maturation and CNS Myelination. Neuron *105*, 293-309.e5.

Yan, F., Al-Kali, A., Zhang, Z., Liu, J., Pang, J., Zhao, N., He, C., Litzow, M.R., and Liu, S. (2018). A dynamic N6-methyladenosine methylome regulates intrinsic and acquired resistance to tyrosine kinase inhibitors. Cell Res *28*, 1062–1076.

Yankova, E., Blackaby, W., Albertella, M., Rak, J., De Braekeleer, E., Tsagkogeorga, G., Pilka, E.S., Aspris, D., Leggate, D., Hendrick, A.G., et al. (2021). Small-molecule inhibition of METTL3 as a strategy against myeloid leukaemia. Nature *593*, 597–601.

Yoon, K.-J., Ringeling, F.R., Vissers, C., Jacob, F., Pokrass, M., Jimenez-Cyrus, D., Su, Y., Kim, N.-S., Zhu, Y., Zheng, L., et al. (2017). Temporal Control of Mammalian Cortical Neurogenesis by m6A Methylation. Cell *171*, 877-889.e17.

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS: A Journal of Integrative Biology *16*, 284–287.

Yu, J., Chen, M., Huang, H., Zhu, J., Song, H., Zhu, J., Park, J., and Ji, S.-J. (2018). Dynamic m6A modification regulates local translation of mRNA in axons. Nucleic Acids Res *46*, 1412–1423.

Yue, Y., Liu, J., Cui, X., Cao, J., Luo, G., Zhang, Z., Cheng, T., Gao, M., Shu, X., Ma, H., et al. (2018). VIRMA mediates preferential m6A mRNA methylation in 3'UTR and near stop codon and associates with alternative polyadenylation. Cell Discov *4*, 10.

Zaccara, S., and Jaffrey, S.R. (2020). A Unified Model for the Function of YTHDF Proteins in Regulating m6A-Modified mRNA. Cell.

Zhang, C., Chen, Y., Sun, B., Wang, L., Yang, Y., Ma, D., Lv, J., Heng, J., Ding, Y., Xue, Y., et al. (2017). m 6 A modulates haematopoietic stem and progenitor cell specification. Nature *549*, 273–276.

Zhang, Z., Zhan, Q., Eckert, M., Zhu, A., Chryplewicz, A., De Jesus, D.F., Ren, D., Kulkarni, R.N., Lengyel, E., He, C., et al. (2019). RADAR: differential analysis of MeRIP-seq data with a random effect model. Genome Biology *20*, 294.

Zhang, Z., Luo, K., Zou, Z., Qiu, M., Tian, J., Sieh, L., Shi, H., Zou, Y., Wang, G., Morrison, J., et al. (2020). Genetic analyses support the contribution of mRNA N 6 -methyladenosine (m 6 A) modification to human disease heritability. Nature Genetics *52*, 939–949.

Zhao, S., Liu, J., Nanga, P., Liu, Y., Cicek, A.E., Knoblauch, N., He, C., Stephens, M., and He, X. (2019). Detailed modeling of positive selection improves detection of cancer driver genes. Nature Communications *10*, 3399.

Zheng, G., Dahl, J.A., Niu, Y., Fedorcsak, P., Huang, C.-M., Li, C.J., Vågbø, C.B., Shi, Y., Wang, W.-L., Song, S.-H., et al. (2013). ALKBH5 Is a Mammalian RNA Demethylase that Impacts RNA Metabolism and Mouse Fertility. Mol Cell *49*, 18–29.

Zhong, S., Li, H., Bodi, Z., Button, J., Vespa, L., Herzog, M., and Fray, R.G. (2008). MTA is an Arabidopsis messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. Plant Cell *20*, 1278–1288.

Zhou, J., Wan, J., Gao, X., Zhang, X., Jaffrey, S.R., and Qian, S.-B. (2015). Dynamic m 6 A mRNA methylation directs translational control of heat shock response. Nature *526*, 591–594.

Zhou, J., Wan, J., Shu, X.E., Mao, Y., Liu, X.-M., Yuan, X., Zhang, X., Hess, M.E., Brüning, J.C., and Qian, S.-B. (2018). N6-Methyladenosine Guides mRNA Alternative Translation during Integrated Stress Response. Mol. Cell *69*, 636-647.e7.

Zhou, K.I., Shi, H., Lyu, R., Wylder, A.C., Matuszek, Ż., Pan, J.N., He, C., Parisien, M., and Pan, T. (2019a). Regulation of Co-transcriptional Pre-mRNA Splicing by m6A through the Low-Complexity Protein hnRNPG. Mol. Cell *76*, 70-81.e9.

Zhou, K.I., Shi, H., Lyu, R., Wylder, A.C., Matuszek, Ż., Pan, J.N., He, C., Parisien, M., and Pan, T. (2019b). Regulation of Co-transcriptional Pre-mRNA Splicing by m6A through the Low-Complexity Protein hnRNPG. Mol. Cell *76*, 70-81.e9.

Zhou, Y., Zeng, P., Li, Y.-H., Zhang, Z., and Cui, Q. (2016). SRAMP: prediction of mammalian N6-methyladenosine (m6A) sites based on sequence-derived features. Nucleic Acids Res. *44*, e91.